
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：贵州师范学院

作品成员：何丽 严亚运 陈飞

指导教师：雍进军

摘要：文章先运用 R 工具对 1999—2013 年广州市的主要财政收入：个人所得税、企业所得税、营业税、增值税等数据进行逐步回归分析，得出预测各税收的回归模型；其次剔除不显著的因素，得到影响财政收入的关键影响因素，通过残差分析来判断模型的拟合度，得到最优模型；最后用指数平滑预测出 2014 和 2015 年广州市财政收入中个人所得税、企业所得税、营业税、增值税的预测值，并结合其他影响财政收入的因素，对广州市 2015 年的财政总收入进行预测；结合社会经济发展和广州市近几年的财政收入及支出等情况，从财政收入和支出预算的角度，向广州市财政局提出几点建议。

关键词：财政收入 回归模型 逐步回归 指数平滑预测

Abstract:Article first use R tools for the main revenue: 1999-2013 in guangzhou city of personal income tax, enterprise income tax, business tax and value-added tax stepwise regression analysis, the data of tax of forecasting regression model; Second shave no significant factors, key factors affecting the parameters that influence the fiscal revenue, judging by residual analysis of the model fitting and get the optimal model; With exponential smoothing finally predicted in 2014 and 2015, the guangzhou fiscal revenue of individual income tax, enterprise income tax, business tax, value-added tax forecast, and combined with other factors that influence the revenue forecast for guangzhou's fiscal revenue in 2015; Combined with social and economic development and guangzhou in recent years, the financial revenue and expenditure, etc., from the perspective of fiscal revenue and expenditure budget, put forward some suggestion to the guangzhou bureau of finance.

Keywords:Fiscal revenueThe regression modelStepwise regressionExponential smoothing forecast

目录

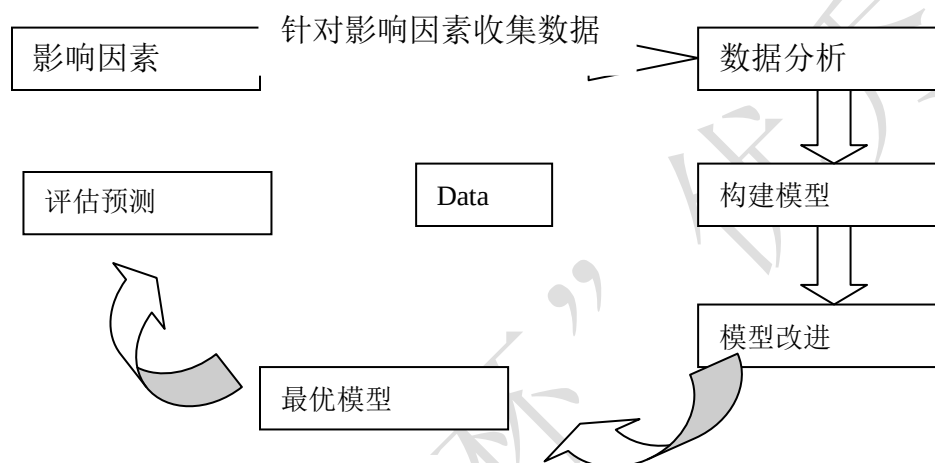
一研究目标·····	01
二分析方法与过程·····	01
(一) 总体流程·····	01
(二) 具体步骤·····	01
1.1 问题 1 的重述·····	01
1.2 问题 1 的分析、预测模型的建立与求解·····	01
(1) 广州市财政收入中个人所得税的预测型·····	02
(2) 广州市财政收入中营业税的预测模型·····	08
(3) 广州市财政收入中企业所得税的预测模型·····	11
(4) 广州市财政收入中个增值税的预测模型·····	13
2.1 问题 2 的重述·····	16
2.2 问题 2 的分析和求解·····	19
3.1 问题 3 的重述·····	19
3.2 问题 3 的分析和求解·····	19
3.2.1 对 2014 年广州市主要财政支出分析	19
3.2.2 建议	21
三、参考文献·····	23

一、研究目标

1. 数据挖掘是指从大量的数据中通过算法搜索隐藏于其中信息的过程，利用广州市往年影响财政收入的因素积累下来的真实数据，采用数据挖掘技术，分析每个因素对财政收入的影响；
2. 构建不同的回归模型，剔除对财政收入影响小的因素，以达到一个最优模型；
3. 预测下一年的财政收入，从而对财政收入支出与收入方向做出备案；
4. 通过对近几年广州市财政收入与支出的数据分析，向广州市财政局提出几点建议。

二、分析方法与过程

（一） 总体流程



（二） 具体步骤

2.1 问题 1 重述

梳理影响广州市财政收入关联指标的有关数据，分析、识别影响财政收入的关键影响因素；

2.2 问题分析、模型的建立与求解

研究表明，影响财政收入的因素有 GDP、固定资产投资、人口总数、就业人数、农业总产量、商业税收（个人所得税、企业所得税、营业税、增值税）等因素有关。税收收入是财政收入的主要来源，我们在此分析影响广州市各税收收入的因素，通过逐步回归分析，得到各种税收及其变量（主要影响因素）的函数关系。

(1) 广州市财政收入中个人所得税的预测模型

步骤 1：数据分析个人所得税

影响个人所得税分因素有城市居民年人均可支配收入、城镇单位职工年平均工资、城镇居民储蓄存款余额、地区生产总值、第二产业增加值、城镇非私营单位从业人员数、地方财政收入等。根据收集的数据用 R 对其建立回归模型，然后再根据建立的模型筛选出相关性大的数据进行逐步回归，直到得到最优模型为止。由得出的 P 值进行判断，若其 P 值小于 0.05 这说明影响显著。

步骤 2：符号说明

y ：个人所得税；

x_1 ：城市居民年人均可支配收入；

x_2 ：城镇单位职工年平均工资；

x_3 ：城镇居民储蓄存款余额；

x_4 ：地区生产总值；

x_5 ：第二产业增加值；

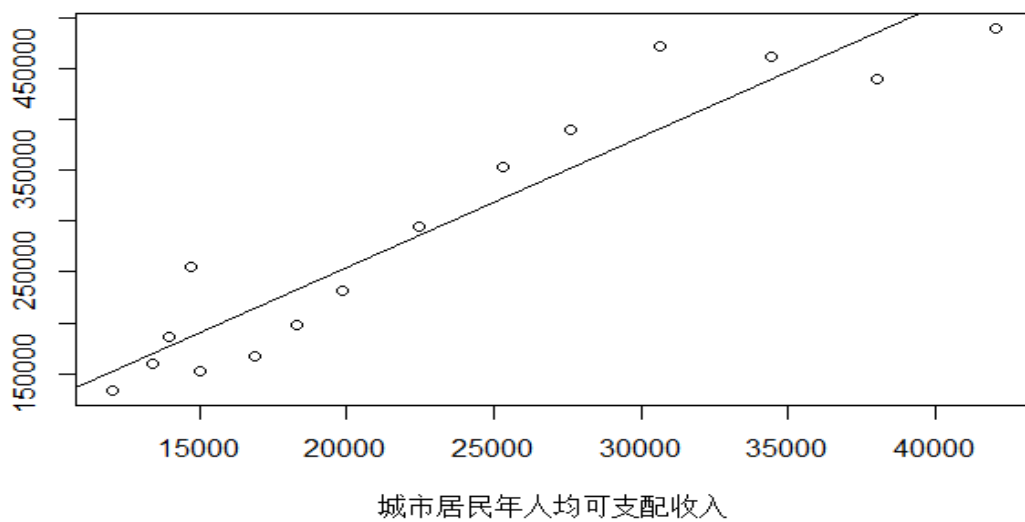
x_6 ：城镇非私营单位从业人员数

x_7 ：地方财政收入

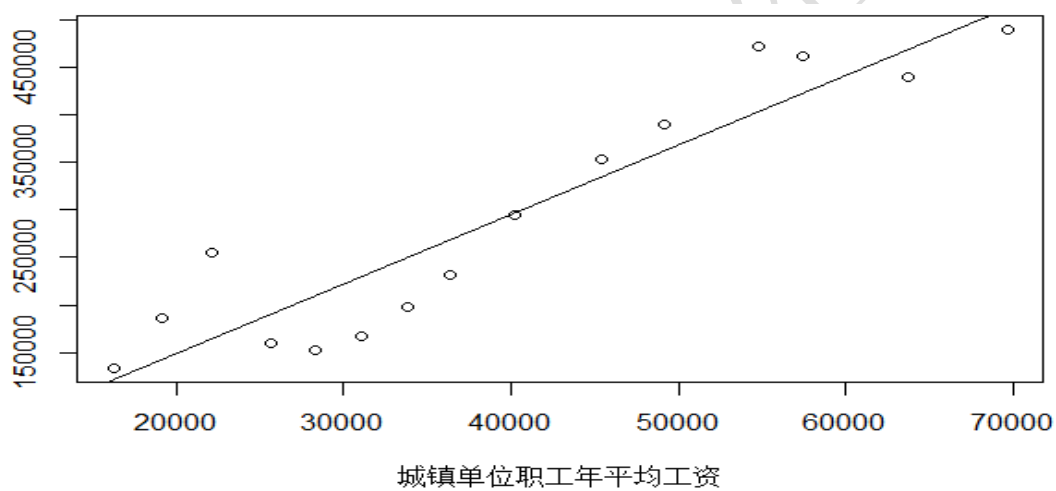
$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$ ：回归系数；

步骤 3：问题的分析

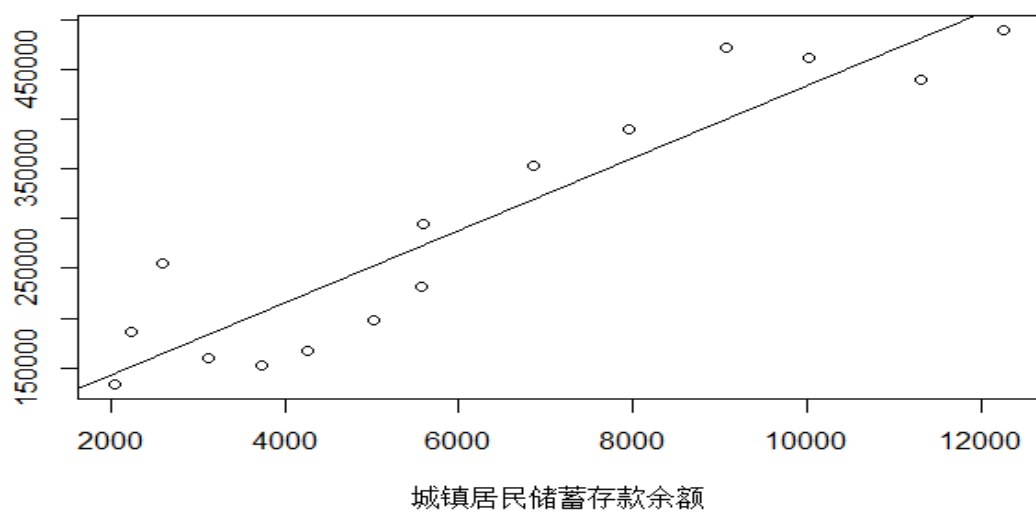
个人所得税

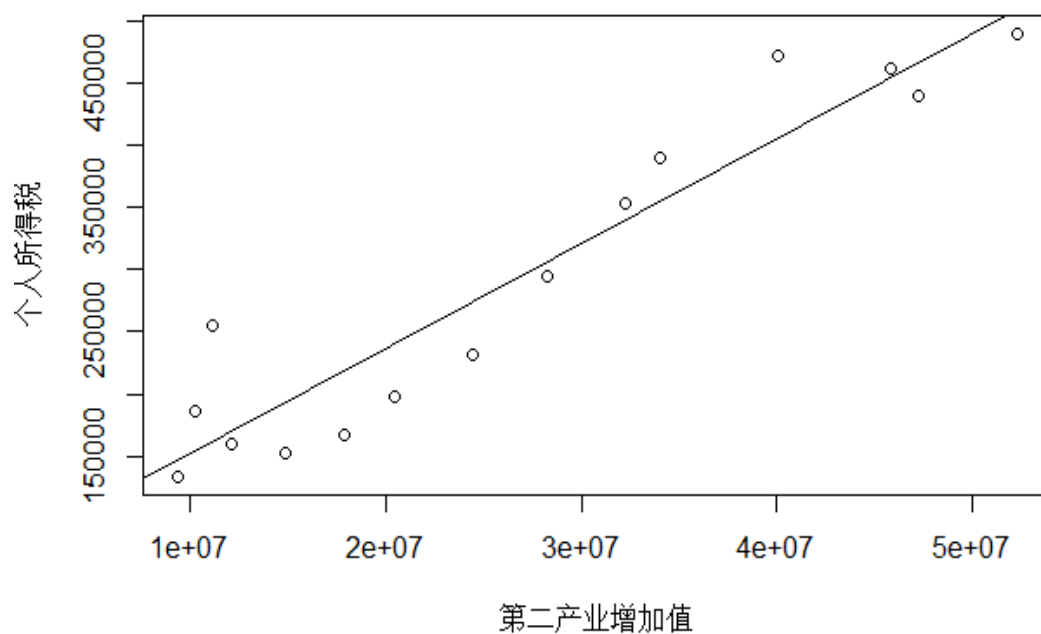
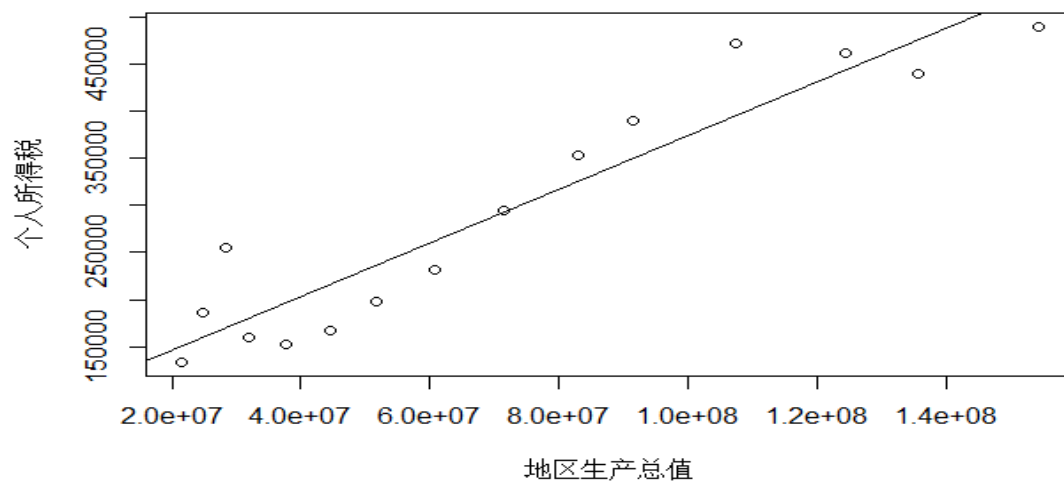


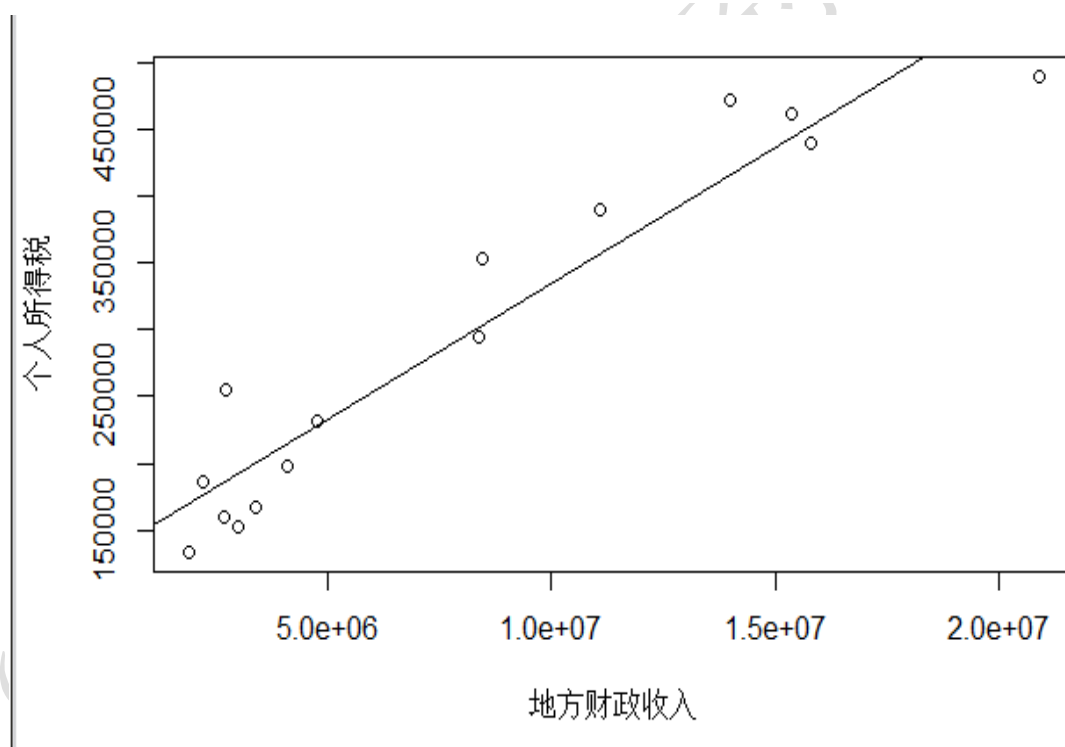
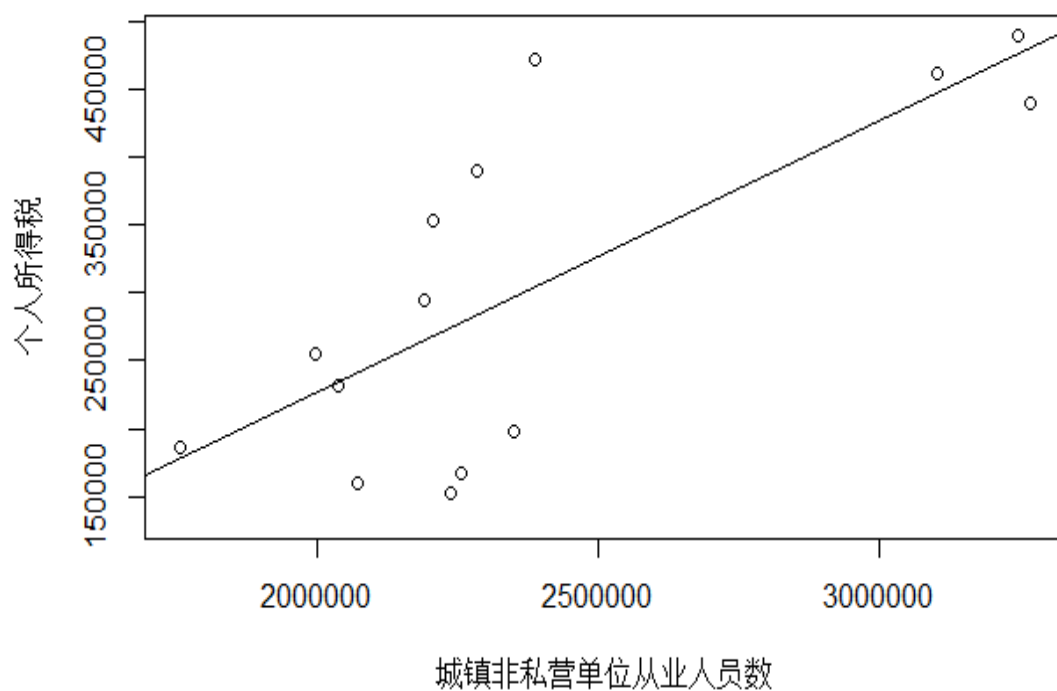
个人所得税



个人所得税







步骤 4：预测模型的建立

对以上的即 y 对 x_1 到 x_6 初步的回归分析，我们再引入一个常量回归系数 β_0 ，作出了初步的模型：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 \quad (1)$$

步骤 5：预测模型的求解

lm.test<-lm(个人所得税~城市居民年人均可支配收入+城镇单位职工年平均工资+城镇居民储蓄存款余额+地区生产总值+第二产业增加值+城镇非私营单位从业人员数+地方财政收入)

summary(lm.test)

confint(lm.test, level=0.95)

得到相关系数：

参数	参数估计值	参数置信区间
β_0	-5.174e+04	-7.933248e+05 6.898438e+05
β_1	3.216e+01	-1.796393e+01 8.228728e+01
β_2	-2.487e+00	-2.875532e+01 2.378231e+01
β_3	3.713e+01	-9.716576e+01 1.714223e+02
β_4	-1.893e-02	-4.598321e-02 8.124123e-03
β_5	2.539e-02	-5.933621e-03 5.671626e-02
β_6	-4.212e-02	-1.989143e-01 1.146823e-01
β_7	3.154e-02	-5.892141e-03 6.896770e-02
R ² =0.925 F= 23.91 p= 0.0005602		

步骤 6：预测模型的改进

由以上的分析，我们剔除 x_2 :城镇单位职工年平均工资 得到模型

$$y = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 \quad (2)$$

lm.test2<-lm(个人所得税~城市居民年人均可支配收入+城镇居民储蓄存款余额+地区生产总值+第二产业增加值+城镇非私营单位从业人员数+地方财政收入)

summary(lm.test2)

confint(lm.test2, level=0.95)

再次检验相关参数

参数	参数估计值	参数置信区间
----	-------	--------

β_0	-1.050e+05	-5.389357e+053.288981e+05
β_1	3.461e+01	-3.995340e+00 7.320962e+01
β_3	2.936e+01	-6.614721e+01 1.248602e+02
β_4	-1.985e-02	-4.252688e-02 2.818814e-03
β_5	2.542e-02	-2.732119e-03 5.356761e-02
β_6	-3.724e-02	-1.703192e-01 9.584398e-02
β_7	3.145e-02	-2.178862e-03 6.507413e-02
R ² = 0.935 F= 32.25 P=9.313e-05		

剔除 x_6

$$y = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_7 x_7 \quad (3)$$

lm.test3<-lm(个人所得税~城市居民年人均可支配收入+城镇居民储蓄存款余额+地区生产总值+第二产业增加值+地方财政收入)

summary(lm.test3)

confint(lm.test2, level=0.95)

再次检验相关参数：

参数	参数估计值	参数置信区间
β_0	-2.121e+05	-5.118882e+05 8.759588e+04
β_1	3.790e+01	2.452034e+00 7.334396e+01
β_3	3.524e+01	-5.359528e+01 1.240809e+02
β_4	-2.339e-02	-4.172278e-02 -5.050066e-03
β_5	3.088e-02	1.035831e-02 5.139368e-02
β_7	3.288e-02	4.855751e-03 6.091309e-02
R ² = 0.940 F= 45.4 P= 4.074e-06		

剔除 x_3

$$y = \beta_0 + \beta_1 x_1 + \beta_4 x_4 + \beta_5 x_5 + \beta_7 x_7 \quad (4)$$

lm.test4<-lm(个人所得税~城市居民年人均可支配收入+地区生产总值+第二产业增加值+地方财政收入)

summary(lm.test4)

confint(lm.test4, level=0.95)

再次检验相关参数：

β_0	-2.121e+05	9.716626e-01 6.812301e+01
β_1	3.790e+01	-3.069477e-02 -5.679556e-03
β_4	3.524e+01	9.347009e-03 4.754421e-02
β_5	-2.339e-02	4.169174e-03 4.869286e-02
β_7	3.088e-02	9.716626e-01 6.812301e+01
R ² = 0.941 F= 57.68 P=7.169e-07		

步骤 7：结果分析

由数据可知，地区生产总值和第二产业增加值对个人所得税的影响最大。

模型（4）的拟合度已经达到了R²=0.941，趋于1，F=57.68趋于最大，P=7.169e-07趋于最小，再通过残差函数对模型的拟合度的检验，确定模型是否达到最优，也可以用step函数看AIC的值在整个逐步回归模型中是否最小。

得到最优预测模型：

$$y = -2.121e+05 + 3.790e+01 x_1 + 3.524e+01 x_4 - 2.339e-02 x_5 + 3.088e-02 x_7$$

(2) 广州市财政收入中营业税的预测模型

步骤 1：数据分析营业税

对营业税进行分析，与营业税相关的因素有公路货运量、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、全社会住宅投资额、地方财政收入、建筑业总产值、住宿和餐饮业零售额、限额以上餐饮业主营业务收入等。根据收集的数据用R对其建立回归模型，然后再根据建立的模型筛选出相关性大的数据进行逐步回归，直到得到最优预测模型为止。由得出的P值进行判断，若其P值小于0.05这说明影响显著。

步骤 2：符号说明

y : 营业税

x_1 : 公路货运量、

x_2 : 公路客运量

x_3 : 建筑业增加值

x_4 : 第三产业增加值

x_5 : 全社会房地产开发投资额

x_6 : 全社会住宅投资额

x_7 : 地方财政收入

x_8 : 建筑业总产值

x_9 : 住宿和餐饮业零售额

x_{10} : 限额以上餐饮业主营业务收入。

步骤 3 预测模型的建立

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10}$$

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}$: 回归系数

步骤 4: 预测模型的求解、改进和结论分析

(1) 预测模型的求解和改进

用 R 软件建立一个 $lm()$ 函数, 对 $lm()$ 函数进行总结, 用 $summary(lm())$ 得到模型 (1) 的回归系数估计值及置信区间 (置信水平 $\alpha=0.05$), 通过逐步回归法剔除对营业税的影响不显著的因素, 用专业的统计分析软件 R 对数据进行逐步回归分析, 最终得到如表所示的结果:

剔除	M-R ²	O-R ²	F	P
原值	0.9961	0.9892	143.4	1.707E-5
x_5	0.9961	0.9892	143.4	1.707E-5
x_{10}	0.9961	0.9910	193.4	1.143E-5
x_9	0.9954	0.9908	216.9	1.209E-7

x_4	0.9936	0.9888	207.1	2.48E-8
x_1	0.9925	0.9884	239.7	2.733E-9

输出结果报告中。列出了用逐步回归法挑选自标量过程，五个变量 $x_5, x_{10}, x_9, x_4, x_1$ ，在 0.05 的水平下是不显著的被依次踢出，同时还概要地提供了每个回归模型变化时的 R^2 值, F 值, 相应的统计量 P 值, $y = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7$

利用 $|Pr| > 0.05$ 的数被踢掉，最后留下的都是小于 0.05 的，说明留下的因素对营业税的影响都是显著，最后得到一个最优预测模型。

(2) 结论分析：

由以上可知，公路客运量、建筑业增加值、全社会住宅投资额、地方财政收入和建筑业总产值对营业税的影响最大。

```
lm.test6<-lm(营业税~公路客运量+建筑业增加值+全社会住宅投资额+地方财政收入+
建筑业总产值)
```

```
summary(lm.test6)##查看回归系数和模型的检验结果
```

```
Call:
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-60067 -20426 -12281   5633 100319
```

```
Coefficients:
```

```
Estimate Std. Error t value Pr(>|t|)
(Intercept)    -5.712e+05  1.594e+05  -3.583  0.00591 **
公路客运量      5.756e+01  6.438e+00   8.940  9.02e-06 ***
建筑业增加值    7.816e-01  8.439e-02   9.262  6.75e-06 ***
全社会住宅投资额 -2.407e-01  3.480e-02  -6.917  6.93e-05 ***
地方财政收入    3.538e-02  1.492e-02   2.371  0.04183 *
建筑业总产值   -1.781e-01  2.665e-02  -6.683  9.03e-05 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 52850 on 9 degrees of freedom
```

```
Multiple R-squared:  0.9925, Adjusted R-squared:  0.9884
```

```
F-statistic: 239.7 on 5 and 9 DF, p-value: 2.733e-09
```

```
 $R^2 = 0.9925$  F=239.7 P = 2.733e-09
```

模型 (6) 的拟合度已经达到了 $R^2=0.9925$ ，趋于 1， $F=57.68$ 趋于最大， $P=7.169e-07$ 趋于最小，再通过残差函数对模型的拟合度的检验，确定模型是否达到最优，也可以用 step 函数看 AIC 的值在整个逐步回归模型中是否最小。

得到最优预测模型： $y = \beta_0 + \beta_2 x_2 + \beta_3 x_3 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8$

即： $y = -5.712e+05 + 5.756e+01x_2 + x_3 + \beta_6x_6 + \beta_7x_7 + \beta_8x_8$

(3) 广州市财政收入中企业所得税的预测模型

步骤 1 数据分析企业所得税

影响企业所得税的因素有第二产业增加值、第三产业增值税、全社会固定资产投资额、城市商品零售价格指数、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、建筑业总产值、建筑业企业利润总额、限额以上连锁店(公司)零售、地方财政总收等。根据收集的数据用 R 对其建立回归模型，然后再根据建立的模型筛选出相关性大的数据进行逐步回归，直到得到最优预测模型为止。由得出的 P 值进行判断，若其 P 值小于 0.05 这说明影响显著。

步骤 2 符号说明

y : 企业所得

x_1 : 第二产业增加值

x_2 : 第三产业增值税

x_3 : 全社会固定资产投资额

x_4 : 规模以上工业企业盈亏相抵后的利润总额

x_5 : 建筑业总产值

x_6 : 建筑业企业利润总额

x_7 : 地方财政总收

步骤 3: 预测模型的建立

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5 + \beta_6x_6 + \beta_7x_7$$

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$: 回归系数

步骤 4: 模型的求解、改进和结论分析

(1) 预测模型的求解和改进

用 R 软件建立一个 $lm()$ 函数，对 $lm()$ 函数进行总结，用 $summary(lm())$ 得到模型 (1)

的回归系数估计值及置信区间（置信水平 $\alpha=0.05$ ），通过逐步回归法剔除对营业税的影响不显著的因素,用专业的统计分析软件 R 对数据进行逐步回归分析，最终得到如表所示的结果：

剔除	M-R ²	O-R ²	F	P
原值	0.9902	0.9788	86.62	1.333E ⁻⁵
x ₅	0.9902	0.9818	117.9	1130E ⁻⁶
x ₃	0.9898	0.9834	135.4	9604E ⁻⁸
x ₇	0.9833	0.9759	132.7	5.08E ⁻⁸

输出结果报告中。列出了用逐步回归法挑选自变量过程，3 个变量 x₅ 和 x₃ 和 x₇，在 0.05 的水平下是不显著的被依次踢出，同时还概要地提供了每个回归模型变化时的 R² 值, F 值, 相应的统计量 P 值，利用 $|Pr| > 0.05$ 的数被踢掉，最后留下的都是 < 0.05 的，说明留下的因素对营业税的影响都是显著，最后得到一个最优模型。

（2）结论分析

由以上可知，第二产业增加值、第三产业增加值、规模以上工业企业盈亏相抵后的利润总额、建筑业企业利润总额对企业所得税影响最大。

`lm.test4<-lm(企业所得税~第二产业增加值+第三产业增加值+规模以上工业企业盈亏相抵后的利润总额+建筑业企业利润总额)`

`> summary(lm.test4)`

Call:

`lm(formula = 企业所得税 ~ 第二产业增加值 + 第三产业增加值 + 规模以上工业企业盈亏相抵后的利润总额 + 建筑业企业利润总额)`

Residuals:

Min	1Q	Median	3Q	Max
-100863	-14161	-682	30421	61269

Coefficients:

Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.214e+05	5.649e+04	2.149 0.06012
第二产业增加值	3.750e-02	1.252e-02	2.996 0.01504
第三产业增加值	-1.725e-02	6.082e-03	-2.836 0.01952*
规模以上工业企业盈亏相抵后的利润总额	-8.108e-02	2.236e-02	-3.625 0.005 **

建筑业企业利润总额 1.925e+00 3.922e-01 4.908 0.000***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 50550 on 9 degrees of freedom

(1 observation deleted due to missingness)

Multiple R-squared: 0.9833, Adjusted R-squared: 0.9759

F-statistic: 132.7 on 4 and 9 DF, p-value: 5.408e-08

$R^2 = 0.9833$, $F = 132.7$ $P = 5.408e-08$

模型（4）的拟合度已经达到了 $R^2=0.9833$ ，趋于 1， $F=132.7$ ，趋于最大， $P=5.408e-08$ 趋于最小，再通过残差函数对模型的拟合度的检验，确定模型是否达到最优，也可以用 step 函数看 AIC 的值在整个逐步回归模型中是否最小。

得到最优预测模型： $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_4 x_4 + \beta_6 x_6$

即 $y = 1.214e+05 + 3.750e-02 x_1 - 1.725e-02 x_2 - 8.108e-02 x_4 + 1.925e+00 x_6$

(4). 广州市财政收入中个增值税的预测模型

步骤 1 数据分增值税

对增值税进行分析，与增值税相关的因素有商品进口总值、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、规模以上国有及国有控股工业企业企业亏损、建筑业总产值、建筑业企业利润总额、批发零售业增加值等。根据收集的数据用 R 对其建立回归模型，然后再根据建立的模型筛选出相关性大的数据进行逐步回归，直到得到最优预测模型为止。由得出的 P 值进行判断，若其 P 值小于 0.05 这说明影响显著，其 P 值越小影响越显著。

步骤 2 符号说明

y : 增值税

x_1 : 商品进口总值

x_2 : 地区生产总值

x_3 : 工业增加值

x_4 : 批发零售业零售额

x_5 : 工业增加值占 GDP

x_6 : 批发零售业增加值

步骤 3: 预测模型的建立

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$$

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$: 回归系数

步骤 4: 预测模型的求解、改进和结论分析

(1) 预测模型的求解和改进

用 R 软件建立一个 `lm()` 函数, 对 `lm()` 函数进行总结, 用 `summary(lm())` 得到模型 (1) 的回归系数估计值及置信区间 (置信水平 $\alpha=0.05$), 通过逐步回归法剔除对营业税的影响不显著的因素, 用专业的统计分析软件 R 对数据进行逐步回归分析, 最终得到如表所示的结果

剔除	M-R ²	O-R ²	F	P
原始	0.9997	0.9995	4550	3.83E-15
x6	0.9997	0.9995	5947	2.20E-16
x3	0.9992	0.9989	3357	< 2.2E-16
x1	0.9995	0.9993	7316	< 2.2E-16

输出结果报告中。列出了用逐步回归法挑选自变量过程, 3 个变量 x_6 和 x_3 和 x_1 , 在 0.05 的水平下是不显著的被依次踢出, 同时还概要地提供了每个回归模型变化时的 R^2 值, F 值, 相应的统计量 P 值, 利用 $|Pr| > 0.05$ 的数被踢掉, 最后留下的都是 < 0.05 的, 说明留下的因素对营业税的影响都是显著, 最后得到一个最优预测模型。

(2) 结论分析

由以上可知, 第二产业增加值、第三产业增加值、规模以上工业企业盈亏相抵后的利润总额、建筑业企业利润总额对企业所得税影响最大。

```
> lm.test5<-lm(增值税~商品进口总值+工业增加值)
```

```
> summary(lm.test5)
```

Call:

```
lm(formula = 增值税 ~ 商品进口总值 + 工业增加值)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-134464	-40531	10643	42553	138130

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-4.564e+03	4.326e+04	-0.106	0.9177
商品进口总值	-1.662e+03	6.699e+02	-2.481	0.0289 *
工业增加值	6.533e-02	9.071e-03	7.202	1.08e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 79420 on 12 degrees of freedom

Multiple R-squared: 0.9844, Adjusted R-squared: 0.9818

F-statistic: 378.9 on 2 and 12 DF, p-value: 1.436e-11

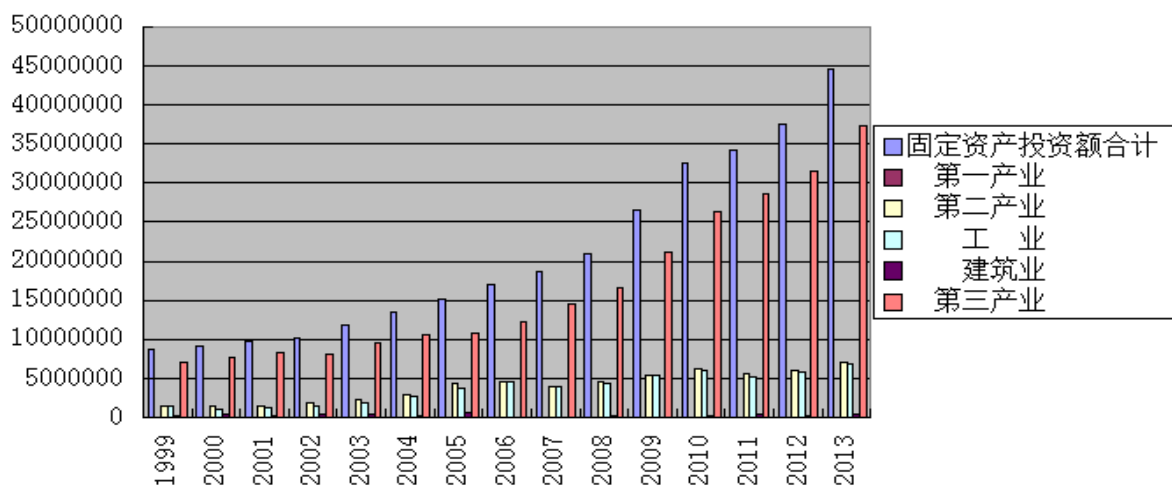
模型(5)的拟合度已经达到了 $R^2=0.9818$, 趋于1, $F=378.9$, 趋于最大, $P=1.436e-11$ 趋于最小, 再通过残差函数对模型的拟合度的检验, 确定模型是否达到最优, 也可以用step函数看AIC的值在整个逐步回归模型中是否最小。

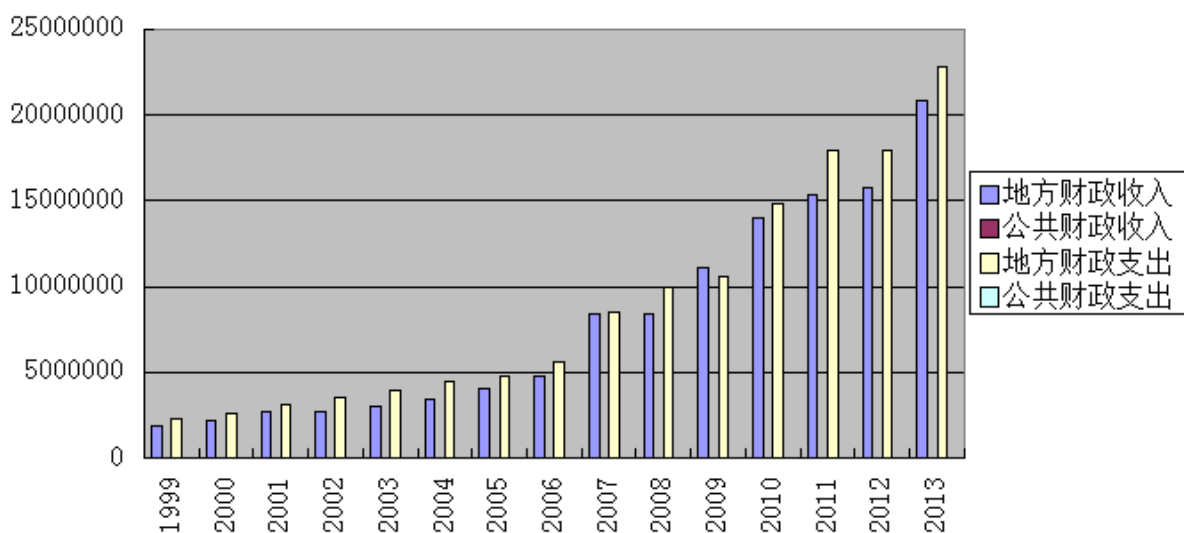
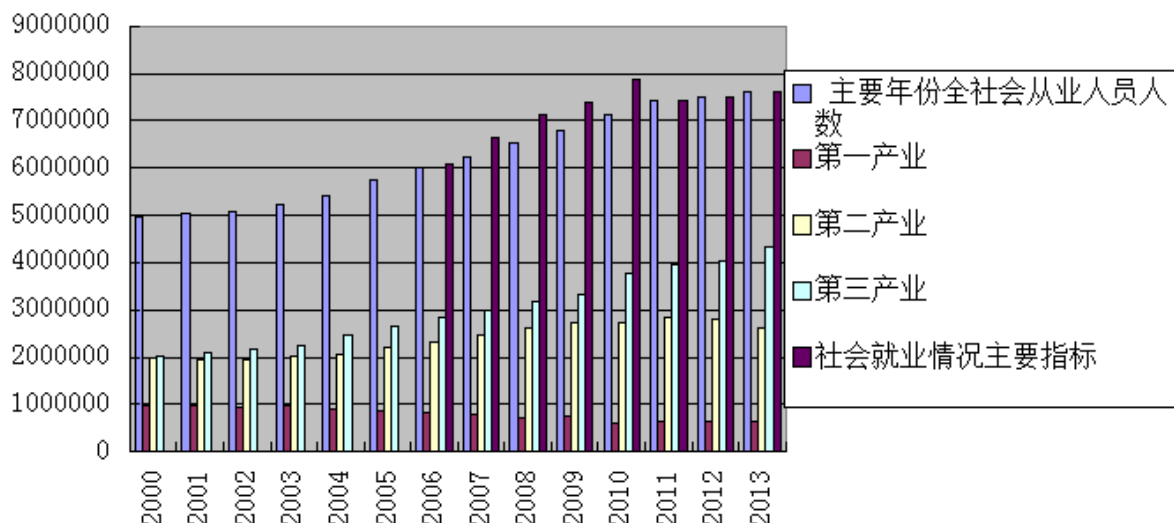
得到最优预测模型:

$$y = \beta_0 + \beta_1 x_1 + \beta_3 x_3$$

$$\text{即 } y = -4.564e+03 - 1.669e+03x_1 + 6.523e-02x_3$$

注: 对财政收入的影响不只是税收影响, 税收是财政收入的主体, 但是还参杂其他的因素。





2.1 问题重述

结合问题 1 的因素分析，利用相关的数据挖掘技术对广州市 2015 年的财政总收入及各个类别收入进行预测；

2.2 问题分析和求解

2.2.1 各个类别收入进行预测

(1) 利用近几年的营业税数据对 2014 到 2015 年的营业税进行预测，用 Holt 指数平滑预测，并将实际值和预测值绘制成的图像进行比较。

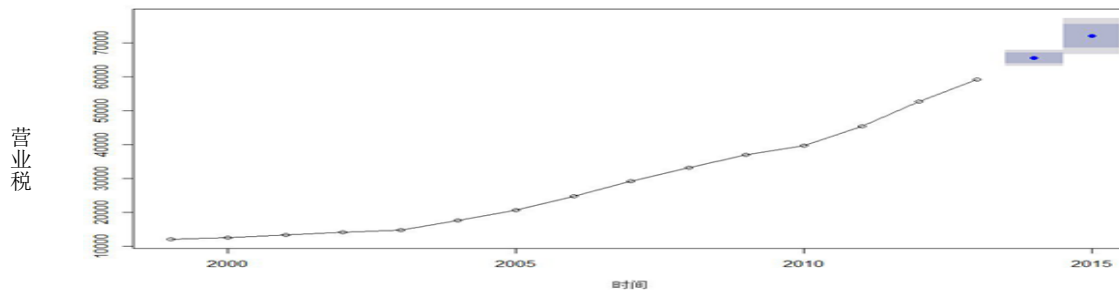
```
financeforecast=forecast.HoltWinters(financeforecast,h=2)> financeforecast
Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2014          65593.53 64027.12 67159.95 63197.91 67989.16
2015          72045.07 68554.02 75536.12 66705.97 77384.17> par(cex=1)>
```

```
plot(financeforecast, type='o', xlab="时间", ylab="营业税", main="")
```

预测结果:

2014 年营业税预测值: 65593.53

2015 年营业税预测值: 72045.07



(2) 利用近几年的企业所得税数据对 2014 到 2015 年的企业所得税进行预测, 用 Holt 指数平滑预测, 并将实际值和预测值绘制成的图像进行比较.

```
financeforecast=forecast.HoltWinters(financeforecast, h=2)
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
--	----------------	-------	-------	-------	-------

2014	56394692	54442806	58346578	53409540	59379845
------	----------	----------	----------	----------	----------

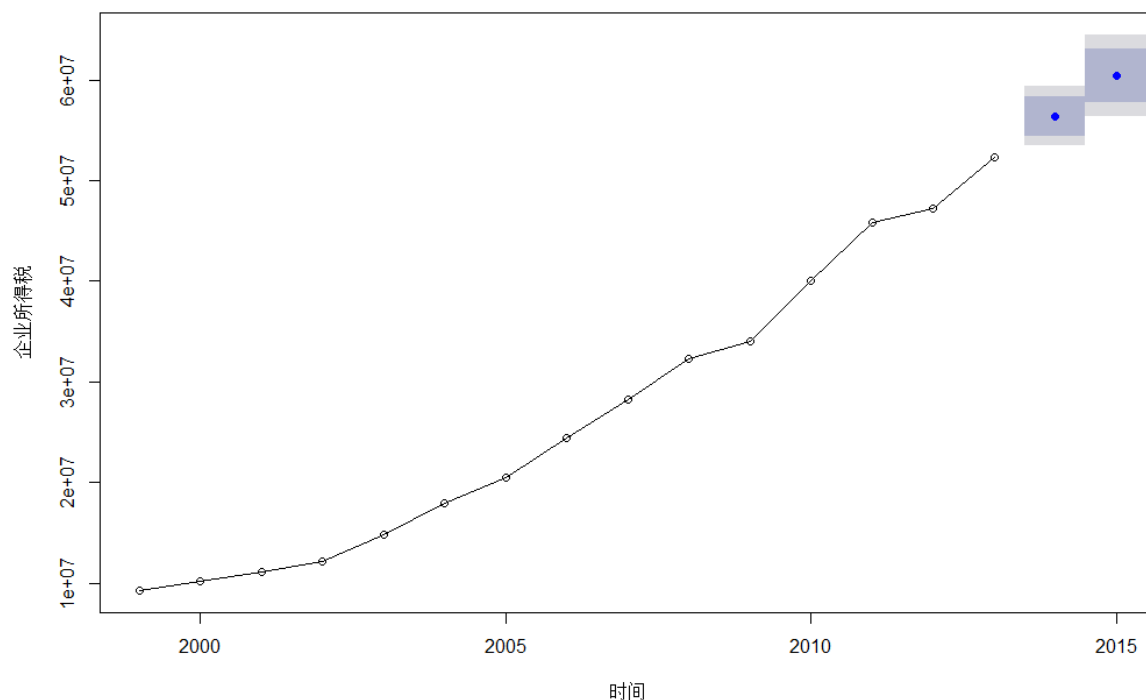
2015	60390481	57753082	63027881	56356927	64424036
------	----------	----------	----------	----------	----------

```
plot(financeforecast, type='o', xlab="时间", ylab="企业所得税", main="")
```

预测结果:

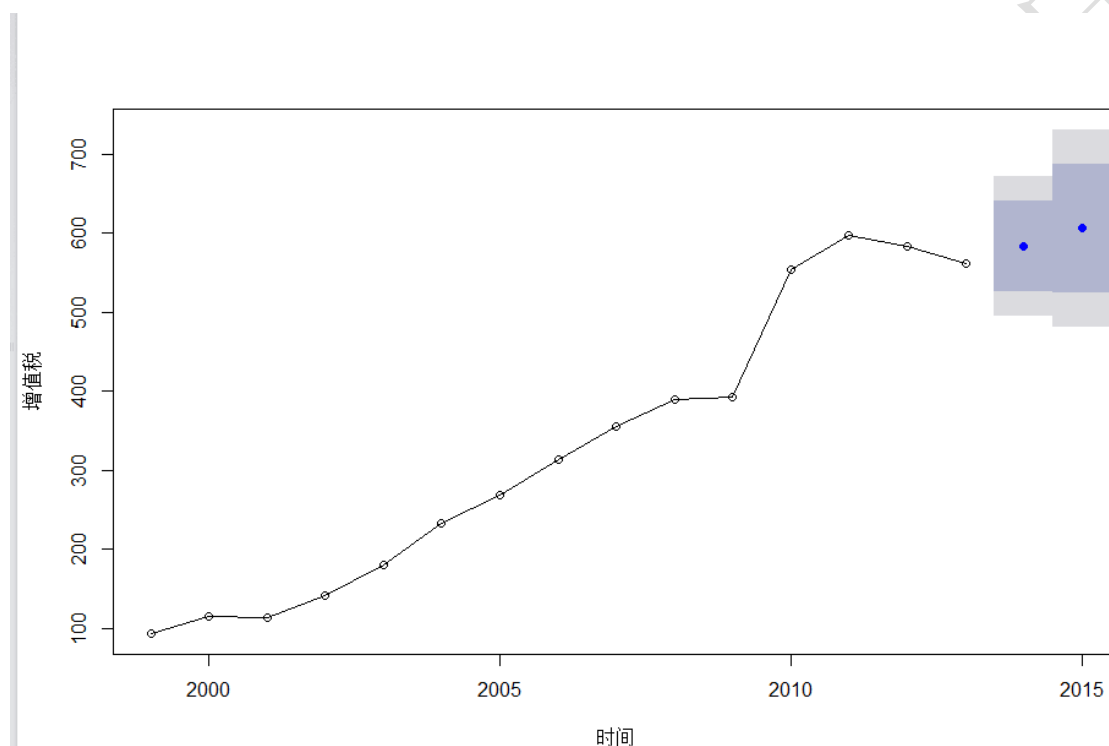
2014 年企业所得税预测值: 56394692

2015 年企业所得税预测值: 60390481



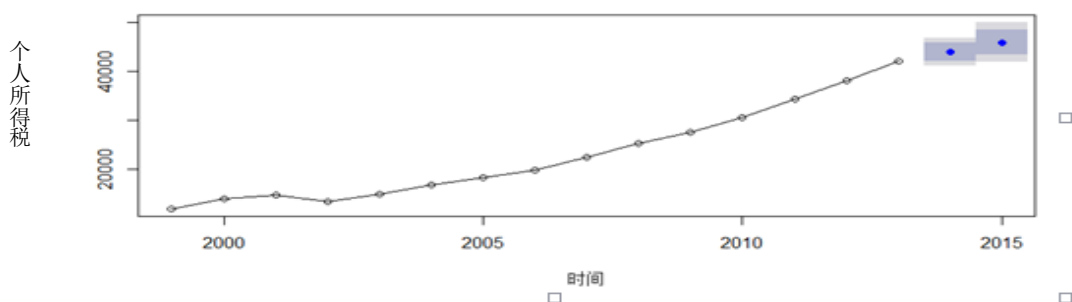
(3) 利用近几年的增值税数据对 2014 到 2015 年的增值税进行预测, 用 Holt 指数平滑预测, 并将实际值和预测值绘制成的图像进行比较.

```
legend(x="topleft", legend=c("观测值", "拟合值"), lty=1:2, cex=1) > financeforec
ast=forecast.HoltWinters(financeforecast, h=2) > financeforecast      Point Fo
recast      Lo 80      Hi 80      Lo 95      Hi 95
2014          583.31 525.4205 641.1995 494.7757 671.8443
2015          605.73 523.8619 687.5981 480.5235 730.9365 > par(cex=1) > plot(f
inanceforecast, type='o', xlab="时间", ylab="增值税", main="")
> plot(financeforecast, type='o', xlab="时间", ylab="增值税", main="")
预测结果:
2014 年增值税预测值: 56394692
2015 年增值税预测值: 60390481
```



(4) 利用近几年的个人所得税数据对 2014 到 2015 年的个人所得税进行预测，用 Holt 指数平滑预测，并将实际值和预测值绘制成的图像进行比较

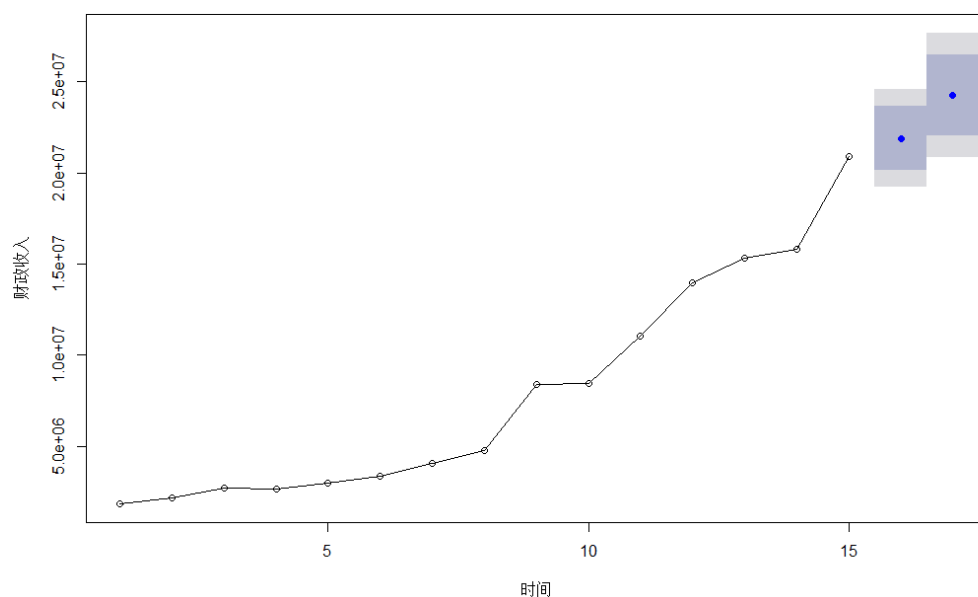
```
financeforecast=forecast.HoltWinters(financeforecast, h=2)
> financeforecast
      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
2014          43997 42146.59 45847.41 41167.04 46826.96
2015          45945 43328.13 48561.87 41942.84 49947.16
> plot(financeforecast, type='o', xlab="时间", ylab="个人所得税", main="")
预测结果:
2014 年个人所得税预测值: 43997
2015 年个人所得税预测值: 45945
```



2.2.2 对广州市 2015 年财政收入的预测

对 2014 年和 2015 年财政总收入的预测值与财政总收入的实际值绘图:

```
plot(financeforecast, type='o', xlab="时间", ylab="财政收入", main="")
```



3.1 问题重述

结合社会经济发展和广州市近几年的财政收入及支出等情况，从财政收入和支出预算的角度，向广州市财政局提出几点建议。

3.2 问题分析和求解

3.2.1 对 2014 年广州市主要财政支出分析

（1）城市基础设施建设 229.8 亿元

去年财政支出 229.8 亿元，用于加大对城市轨道、铁路、公路等交通基础设施和农田水利设施、环境保护设施的投入，加快增强国家中心城市带动和辐射功能的交通枢纽工程建设，以财政资金投入带动社会固定资产投资平稳增长。

（2）商业服务业 14.9 亿元

2014 年商业服务业支出 14.9 亿元，鼓励外资总部企业落户我市，支持外贸企业做强做大出口市场，重点支持外贸新业态培育、商业模式创新和市场拓展。

（3）金融业 2.7 亿元

去年金融支出 2.7 亿元，支持金融企业自主创新、技术改造、品牌创建、提升总部功能等生产经营活动，促进广州金融业发展和区域金融中心建设。

（4）价格调节基金 5 亿元

市财政支出价格调节基金 5 亿元支持重点商品储备，调控价格，稳定市场，发放政策性补贴，为保障供给、促进流通和结构调整进行政府资助对低收入群体实施救助等。

（5）战略性主导产业 40 亿元

市财政支出 40 亿元支持产业转型升级，重点支持新一代信息技术、生物与健康、电子商务、金融保险、节能与新能源等战略性主导产业发展，支持总部经济和民营经济发展，推进企业改革发展和产业转型升级。

（6）扶持“营改增”试点企业 2 亿元

2014 年全市纳入试点的纳税人约 15 万户，包括交通运输业、邮政业、电信业以及 7 个现代服务业。2014 年向税负增加的“营改增”试点企业拨付 2013 年度财政扶持资金逾共 2 亿元。停征 13 项涉企行政事业性收费，全面下调堤围防护费收费标准，为企业减负约 20 亿元。

（7）教育类 227.7 亿元

2014 年广州教育支出 227.7 亿元，重点用于促进义务教育均衡发展，大力支持学前教育、特殊教育和民办教育，实施从学前教育到高等教育家庭经济困难学生助学资助体系和农村义务教育营养改善计划，促进教育公平。

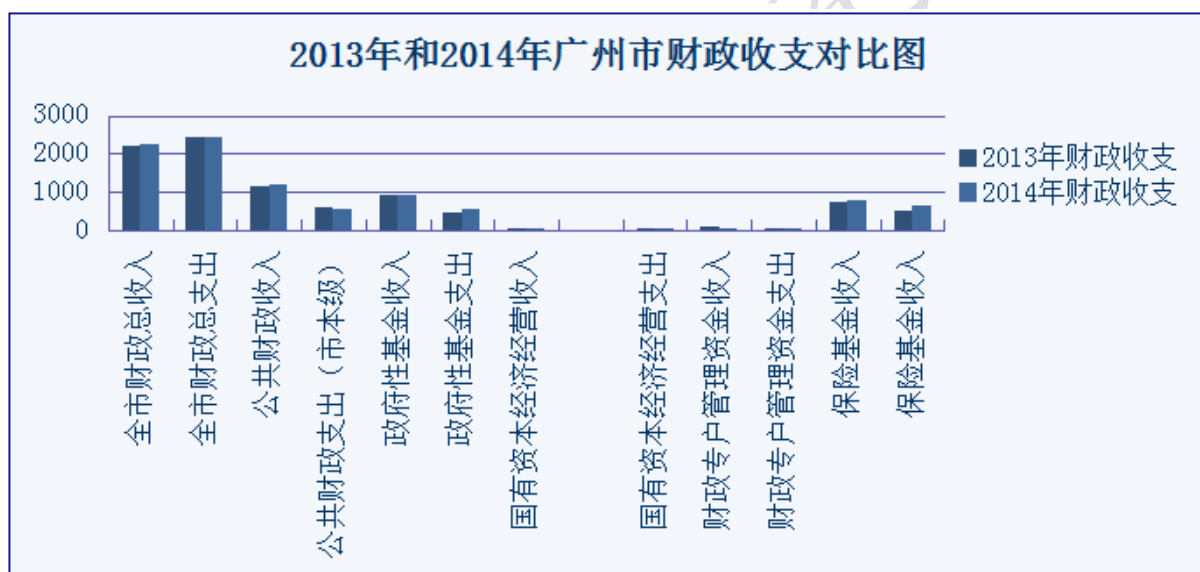
(8) 医疗卫生 115.7 亿元

2014 年医疗卫生支出 115.7 亿元，重点用于基本公共卫生服务支出，为城乡居民免费提供 11 大项、32 小项基本公共卫生服务项目，提高困难群众医疗救助水平。

(9) 社会保障 148.8 亿元

社会保障支出 148.8 亿元，对全市城镇低保标准统一提高到 600 元，农村低保平均标准 577.14 元；提高全市优抚对象补贴生活困难补助标准。

3.2.2 建议



结合上图结果和广州市当前经济情况，对广州市财政局增加财政收入提出如下相关政策建议：

(1) 提高经济发展水平，以经济增长带动财政收入增长。支持经济发展方式转变：安排产业引导基金；大力支持总部企业、民营及中小企业发展；支持新一代信息技术、生物与健康、时尚创新、商贸流通、金融保险、先进制造业等战略性主导产业发展等。无论是长期还是短期，经济发展水平对财政收入的增加都具有显著影响，因此广州市政府应该大力发展生产力，才能通过增加国内生产总值，为财政收入增长提

供的财源。

(2) 优化产业结构，构建现代产业体系。产业结构优化升级与经济的持续发展具有强相关性，加快结构调整步伐，坚持产业高端发展，全力建设现代产业体系。广州市加快第三产业发展，尤其是现代物流、金融保险、商贸会展和文化旅游等税收量高的现代服务业蓬勃发展，积极培育新的财源增长点。

(3) 保持税收收入占 GDP 的合理比例，提高税收征管执行力度。税收收入是广州市财政收入的主要来源，且税收水平对财政收入的冲击很大，因此收税征收部门应该提高税收征管执行力度。税务机关要加强税源监控，强化税收业务知识培训，转变工作作风，依法治税、依法征税，通过加强各部门配合，不断提高税收征收率，保持税收随着经济的发展平稳增长。

(4) 实现充分就业，促进经济快速增长。促进就业作为中央和地方经济工作会议的重要议题，其重要性可想而知。广州市政府相关部门应优化就业环境，建立公平竞争的劳动力市场，引进高端技术人才，开展各种类型的教育培训，提高各层次就业人员的职业素质和专业技能，加大投入发展职业教育落实，从而增加财政收入。

参考文献

- [1] 汪晓银, 周保平主编. 数学建模与数学实验[M]. 北京: 科学出版社, 2012
- [2] <http://www.gzstats.gov.cn/>广州市统计信息网
- [3] 贾俊平著. 统计学—基于 R[M]. 北京: 清华大学出版社, 2014

附件 （见压缩包）