

云南省烟草商业系统卷烟销量影响因素分析

刘忠华

中国烟草总公司云南省公司，云南 昆明

收稿日期：2021年8月30日；录用日期：2021年9月23日；发布日期：2021年9月30日

摘要

本文以云南省烟草商业系统卷烟销量为研究对象，利用2008~2018年的数据，构建基于主成分分析和多元线性回归的统计模型，定量分析云南省地区生产总值、社会消费品零售总额、居民可支配收入、人口、单箱销售收入和税利等因素对卷烟销量的影响，为云南省卷烟经济调控提供决策依据。

关键词

主成分分析，多元线性回归，卷烟销量，影响因素

Analysis on Influencing Factors of Cigarette Sales in Yunnan Tobacco Commercial System

Zhonghua Liu

China National Tobacco Corporation Yunnan Company, Kunming Yunnan

Received: Aug. 30th, 2021; accepted: Sep. 23rd, 2021; published: Sep. 30th, 2021

Abstract

Taking the cigarette sales of Yunnan tobacco commercial system as the research object, using the data from 2008 to 2018, this paper constructs a statistical model based on principal component analysis and multiple linear regression to quantitatively analyze the impact of regional GDP, total retail sales of social consumer goods, residents' disposable income, population, single box sales revenue and tax profit on cigarette sales in Yunnan Province. It provides decision-making basis for cigarette economic regulation in Yunnan Province.

Keywords

Principal Component Analysis, Multiple Linear Regression, Cigarette Sales, Influencing Factors

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

烟草业作为国民经济体系中的一个相对特殊的行业，从短期来看，它的发展对经济是有利的，但从长远来看，它对人民的健康和环境都会产生不利影响，进而造成经济发展的滞后，因此，烟草业受到政府的严格监督与控制。

烟草业是云南省的主要经济产业，早期卷烟产销量的稳定增长为云南省经济发展做出了巨大贡献，但 2008 年以来，如图 1 所示，云南省烟草商业系统卷烟销量增速整体出现放缓趋势，卷烟消费市场开始进入饱和阶段，同时人们注重健康的意识在逐渐增强，卷烟的需求量继而出现减少趋势，另外，我国 2011 年颁布了禁烟通知，也对卷烟销售产生了一定影响。在卷烟消费进入这样一个特殊阶段后，对其影响因素做定量分析有助于全省烟草业的稳定发展。

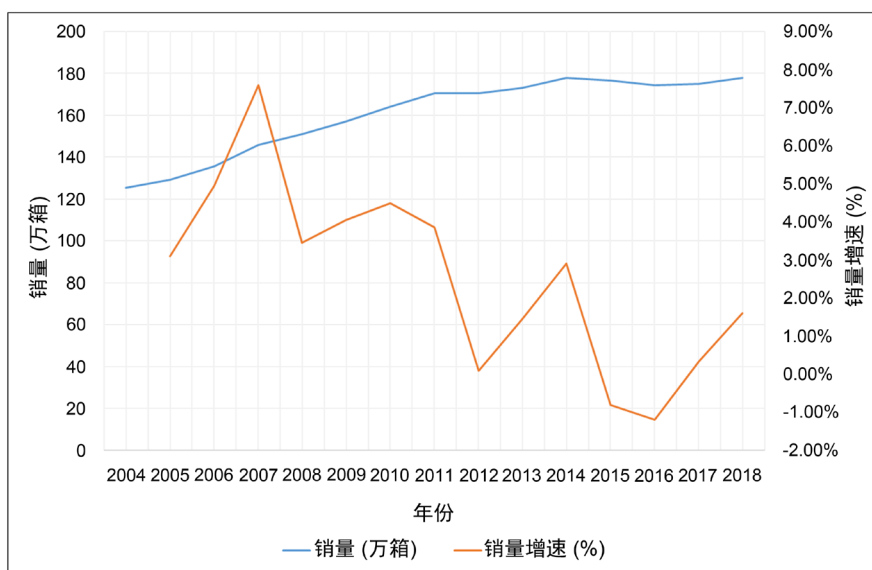


Figure 1. Curve: Change trend of cigarette sales in Yunnan tobacco commercial system from 2004 to 2018

图 1. 2004~2018 年云南省烟草商业系统卷烟销量变化趋势

2. 数据的选取与处理

2.1. 数据选取

卷烟销售量的影响因素来自于地区经济发展水平、相关政策、物价水平等方面，本文从以上角度出发，选取了云南省地区生产总值(简称：GDP)、第三产业增加值占 GDP 的比重、社会消费品零售总额、居民人均可支配收入、城镇居民人均可支配收入、农村居民家庭人均纯收入、居民消费价格指数、年末

常住人口、城镇人口、乡村人口、单箱销售收入、单箱税利、烟草商业系统卷烟销量年度数据作为分析数据，数据来源为国家统计局、《中国烟草年鉴》和《云南年鉴》，样本区间为 2008~2018 年。

2.2. 数据处理

2.2.1. 定基期居民消费价格指数计算

由于与经济金额有关的时间序列数据包含了通货膨胀或紧缩的因素，而度量通货膨胀是居民消费价格指数的基本功能之一，根据国家统计局公布的云南省“居民消费价格指数(上年 = 100)”(简称：CPI(上年 = 100))，计算出以 2008 年为基期的居民消费价格指数：“居民消费价格指数(2008 年 = 100)”(简称：CPI(2008 年 = 100))，具体算法为：

$$2008 \text{ 年 CPI (2008 年 = 100)} = 100,$$

$$2009 \text{ 年 CPI (2008 年 = 100)} = 2009 \text{ 年 CPI (上年 = 100)},$$

$$2010 \text{ 年 CPI (2008 年 = 100)} = 2010 \text{ 年 CPI (上年 = 100)} \times 2009 \text{ 年 CPI (2008 年 = 100)} / 100,$$

$$2011 \text{ 年 CPI (2008 年 = 100)} = 2011 \text{ 年 CPI (上年 = 100)} \times 2010 \text{ 年 CPI (2008 年 = 100)} / 100,$$

$$2012 \text{ 年 CPI (2008 年 = 100)} = 2012 \text{ 年 CPI (上年 = 100)} \times 2011 \text{ 年 CPI (2008 年 = 100)} / 100,$$

……(以此类推) [1]

这样，就得到了各年度具有可比性的通货膨胀度量指标。

2.2.2. 居民人均可支配收入缺失数据测算

国家统计局自 2013 年开始公布居民人均可支配收入，对于 2008~2012 年的居民人均可支配收入，本文通过相关指标测算得到，测算公式为：居民人均可支配收入 = 城镇人口/年末常住人口 × 城镇居民人均可支配收入 + 乡村人口/年末常住人口 × 农村居民家庭人均纯收入。

2.2.3. 单箱税利与销售利税率

为体现单箱税利与单箱销售收入的关系，将单箱税利指标替换为销售利税率，计算公式为：销售利税率 = 单箱税利/单箱销售收入。

2.2.4. 数据处理结果

数据处理完成后，选取的 9 项指标分别为： X_1 —GDP(亿元)、 X_2 —第三产业增加值占 GDP 的比重(%)、 X_3 —社会消费品零售总额(亿元)、 X_4 —居民人均可支配收入(元)、 X_5 —年末常住人口(万人)、 X_6 —单箱销售收入(元)、 X_7 —销售利税率(%)、 X_8 —CPI(2008 年 = 100)、 Y —销量(万箱)，各指标数据见表 1。

Table 1. Data on cigarette sales of tobacco commercial system in Yunnan Province and its influencing factors from 2008 to 2018
表 1. 2008~2018 年云南省烟草商业系统卷烟销量及其影响因素数据

年份	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	Y
2008	6016.6	42.19327	1764.7	6451	4543	14,538	18.54873	100.0	150.93
2009	6574.4	44.40405	2051.1	7127	4571	15,830	19.18789	100.4	157.04
2010	7735.3	43.66088	2555.8	8155	4602	18,709	19.7071	104.1	164.09
2011	9523.1	45.31193	3105.9	9820	4631	22,045	22.95569	109.2	170.42
2012	11,097.4	45.04298	3597.9	11,571	4659	24,163	22.13137	112.2	170.56
2013	12,825.5	46.36466	4112.6	12,578	4687	25,909	20.08433	115.6	173.01
2014	14,041.7	47.41377	4632.9	13,772	4714	27,489	20.43235	118.4	178.04

Continued

2015	14,960.0	49.39171	5103.2	15,223	4742	28,883	23.72363	120.7	176.59
2016	16,369.0	51.29269	5722.9	16,720	4771	28,996	24.69063	122.5	174.46
2017	18,486.0	53.17429	6423.1	18,348	4801	30,000	24.74341	123.6	175.05
2018	20,880.6	53.22883	9197.3	20,084	4830	31,000	24.91366	125.6	177.84

3. 变量相关性

选定变量后, 绘制各自变量的相关系数密度图, 如图 2 所示, 可以看出相关系数基本都在 0.85 以上, 存在显著相关性, 直接用这些数据做回归分析会出现多重共线性问题, 且本文选用的数据样本量较少, 而卷烟销量的影响因素较多, 因此适合用主成分回归方法进行分析。

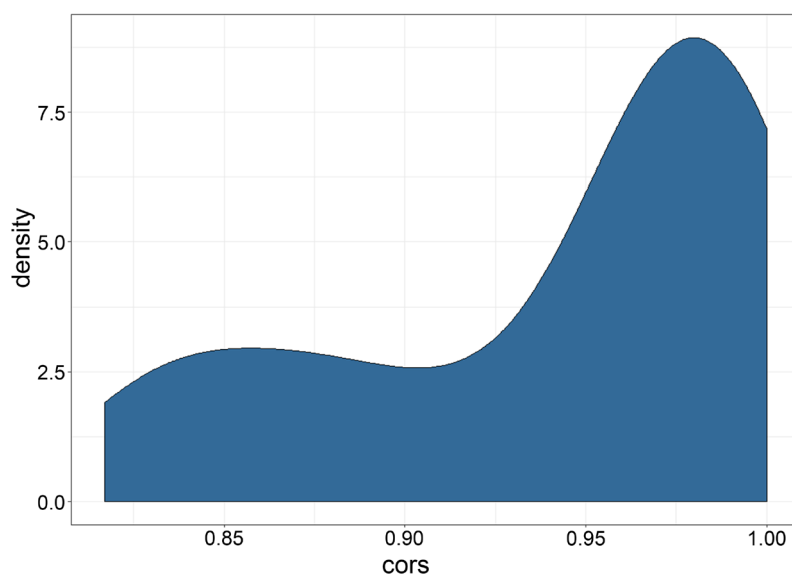


Figure 2. Curve: Correlation coefficient density diagram of independent variables
图 2. 自变量相关系数密度图

4. 主成分回归

主成分回归可以用于解决多重共线性的问题, 主要思路是: 利用主成分分析方法得到互不相关的几个主成分, 再用主成分与因变量建立多元线性回归模型, 由于各主成分是标准化的原始变量的线性组合, 可以将主成分还原为原始变量, 最终就可得到因变量与各自变量的多元线性回归方程[2] [3]。

4.1. 主成分分析

主成分分析结果如图 3 所示, 可以得到各主成分与各标准化变量的关系表达式, 例如, 第一主成分:

$$F_1 = 0.362ZX_1 + 0.354ZX_2 + 0.35ZX_3 + 0.363ZX_4 + 0.363ZX_5 + 0.352ZX_6 + 0.324ZX_7 + 0.358ZX_8$$

其中, ZX_n 为标准化的 X_n , 标准化的规则为: $ZX_n = \frac{X_n - \bar{X}_n}{S_{X_n}}$ ($n=1, 2, \dots, 8$), \bar{X}_n 表示自变量 X_n 的均值,

S_{X_n} 表示 X_n 的标准差。

将标准化的原始变量带入主成分表达式, 可以得到样本数据的各主成分得分情况, 见表 2。

```
> summary(PCA, loadings=TRUE)
Importance of components:
              Comp.1      Comp.2      Comp.3      Comp.4      Comp.5      Comp.6      Comp.7      Comp.8
Standard deviation 2.7452206 0.49658527 0.39561466 0.229473532 0.0673157397 0.0461878404 0.0326020940 1.643429e-02
Proportion of Variance 0.9420295 0.03082462 0.01956387 0.006582263 0.0005664261 0.0002666646 0.0001328621 3.376074e-05
Cumulative Proportion 0.9420295 0.97285415 0.99241802 0.999000287 0.9995667126 0.9998333772 0.9999662393 1.000000e+00

Loadings:
              Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
ZX1 GDP (亿元)      0.362 0.182 0.134      0.366 0.190 0.589 0.547
ZX2 第三产业增加值占GDP的比重(%) 0.354 -0.129 0.351 -0.748 -0.259 -0.269 0.150 -0.115
ZX3 社会消费品零售总额(亿元) 0.350 0.135 0.579 0.604 -0.262 -0.267      -0.117
ZX4 居民人均可支配收入(元) 0.363 0.101      0.500 0.409 -0.110 -0.648
ZX5 年末常住人口(万人) 0.363 0.113      -0.119 -0.272 0.467 -0.613 0.417
ZX6 单箱销售收入(元) 0.352 0.216 -0.582 0.117 -0.516      0.370 -0.255
ZX7 销售利税率(%) 0.324 -0.906 -0.157 0.207
ZX8 CPI (2008年=100) 0.358 0.203 -0.391      0.371 -0.653 -0.316 0.113
```

Figure 3. Curve: Importance of components and loadings

图 3. 主成分贡献率与载荷

Table 2. Score of principal component

表 2. 主成分得分表

年份	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
2008	-4.207	0.063	0.376	0.032	0.158	-0.034	-0.002	0.005
2009	-3.545	-0.116	0.491	-0.293	-0.086	-0.007	0.018	-0.026
2010	-2.826	0.043	0.102	0.061	-0.084	0.062	-0.057	0.022
2011	-1.305	-0.867	-0.311	0.192	-0.051	-0.053	0.041	0.022
2012	-0.731	-0.215	-0.427	0.310	0.052	0.087	0.008	-0.022
2013	-0.223	0.851	-0.293	0.000	-0.006	-0.008	0.026	0.006
2014	0.529	0.945	-0.311	-0.044	-0.019	-0.028	0.000	0.004
2015	1.747	-0.180	-0.416	-0.038	-0.024	-0.038	-0.012	-0.022
2016	2.601	-0.416	-0.159	-0.204	0.045	-0.035	-0.063	0.000
2017	3.437	-0.236	0.141	-0.402	0.038	0.066	0.039	0.015
2018	4.522	0.129	0.806	0.387	-0.023	-0.013	0.002	-0.003

4.2. 多元线性回归

为更好地拟合模型，避免有效信息的丢失，初步选取所有主成分得分与卷烟销量做多元线性回归，再根据显著性逐步剔除不显著的主成分，最终得到主成分回归结果，见图 4。

从主成分回归结果可以看出，回归模型调整后的决定系数为 0.9803，拟合效果较好。常数项和主成分 F_1 、 F_2 、 F_3 、 F_4 、 F_5 的 t 检验 p 值均小于显著性水平 0.05，说明其对因变量的影响是显著的。F 检验的 p 值小于显著性水平 0.05，回归模型显著。

下面针对回归模型残差的正态性、独立性、同方差性以及模型的线性进行诊断：

多元线性回归的正态性假设认为：当预测变量值固定时，因变量呈正态分布，则残差值也应该是一个均值为 0 的正态分布。从图 5 学生化残差 Q-Q 图可以看出，所有点均在置信区间内，表明符合正态性假设。

使用 Durbin-Watson 检验函数检验残差的序列相关性，检验结果如图 6 所示，p 值大于显著性水平 0.05，不显著，说明误差项之间无自相关性，相互独立。

```
> summary(lm.pca)

call:
lm(formula = Y ~ F1 + F2 + F3 + F4 + F5, data = model_data_PCA1)

Residuals:
    2008     2009     2010     2011     2012     2013     2014     2015     2016     2017     2018 
-0.01960 -0.28505 -0.26288  1.20493 -0.13820 -0.87247  1.58093 -1.48999 -0.37255  0.72367 -0.06878 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 169.8209    0.3759  451.830 1.01e-12 ***
F1           2.6710    0.1369  19.509 6.53e-06 ***
F2           2.3975    0.7569   3.168 0.024882 *
F3          -8.3051    0.9500  -8.742 0.000325 ***
F4           4.4242    1.6379   2.701 0.042725 *
F5          -30.3223    5.5834  -5.431 0.002870 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.247 on 5 degrees of freedom
Multiple R-squared:  0.9902,    Adjusted R-squared:  0.9803 
F-statistic: 100.8 on 5 and 5 DF,  p-value: 5.145e-05
```

Figure 4. Curve: Result of principal components regression

图 4. 主成分回归结果

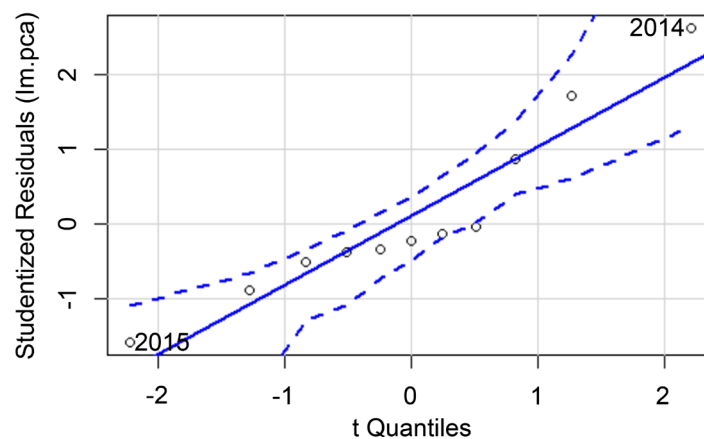


Figure 5. Curve: Q-Q plot of studentized residual

图 5. 学生化残差 Q-Q 图

```
lag Autocorrelation D-W Statistic p-value
 1      -0.4866904      2.972722    0.88
Alternative hypothesis: rho != 0
```

Figure 6. Curve: Result of Durbin-Watson test

图 6. Durbin-Watson 检验结果

进一步通过 NCV 检验对模型残差的异方差性进行检验，NCV 检验零假设为误差方差不变，备择假设为误差方差随着拟合值水平的变化而变化。检验结果如图 7 所示，p 值为 0.16454，大于显著性水平 0.05，不能拒绝误差方差不变的假设，因此推断模型残差的方差是相同的。

```
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 1.931977, Df = 1, p = 0.16454
```

Figure 7. Curve: Result of NCV test

图 7. NCV 检验结果

从图 8 成分残差图可以看出，各主成分与卷烟销量满足线性假设，线性模型对本次分析是适合的。

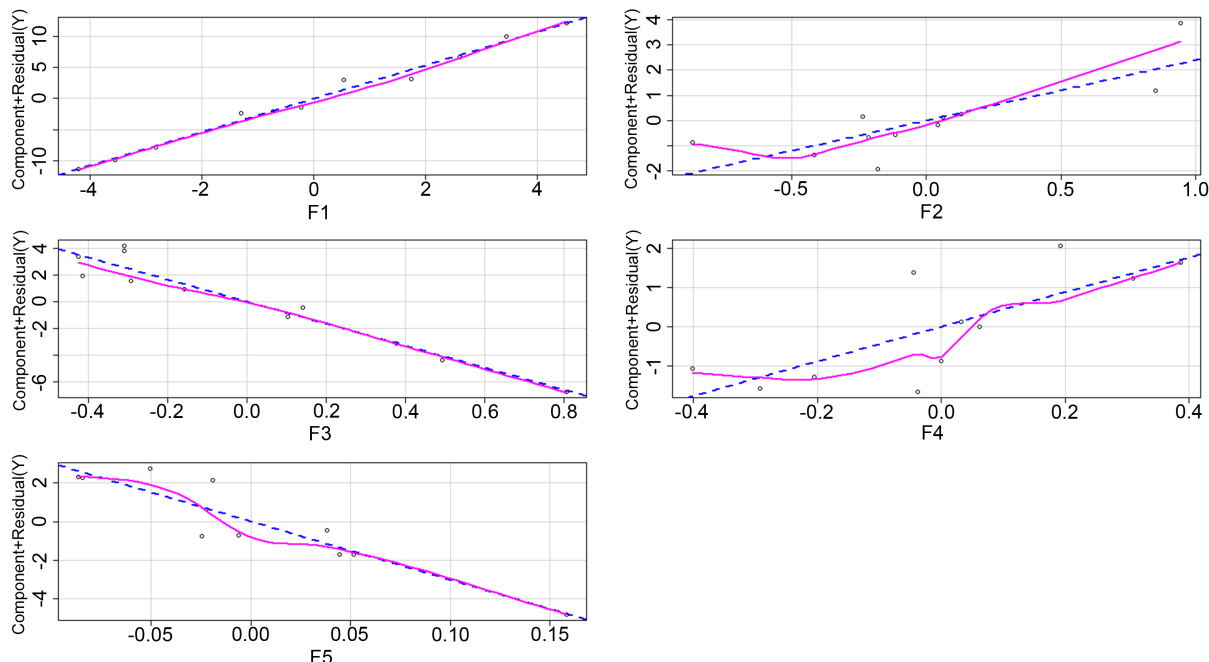


Figure 8. Curve: Component and residual plots
图 8. 成分残差图

最终可以确定云南省烟草商业系统卷烟销量与各主成分的回归模型为：

$$Y = 169.8209 + 2.671F_1 + 2.3975F_2 - 8.3051F_3 + 4.4242F_4 - 30.3223F_5$$

将主成分还原为标准化的原始变量，得到模型表达式为：

$$Y = 169.8209 - 10.6226ZX_1 + 2.2708ZX_2 + 7.0546ZX_3 - 14.8145ZX_4 + 9.1519ZX_5 + 22.4749ZX_6 - 0.7089ZX_7 - 6.6527ZX_8$$

最后把标准化的原始变量转化为原始变量，可以得到烟草商业系统卷烟销量与各原始变量的回归方程：

$$Y = -284.8592 - 0.0023X_1 + 0.6202X_2 + 0.0034X_3 - 0.0034X_4 + 0.1014X_5 + 0.0041X_6 - 0.3079X_7 - 0.7487X_8$$

根据最终的主成分回归模型，得到拟合值与真实值的对比图如图 9 所示，可以看出，模型拟合值和实际销量的整体趋势是一致的。

5. 模型结果分析

从最终的回归方程可以看出：

常数项为负，当各项指标都为零时，模型拟合结果为负值，表明这种极端情况是不足以支撑云南省烟草商业系统产生卷烟销量的，当各项指标增长到一定程度后，卷烟销量才开始为正。

X_1 GDP (亿元)的系数为负，表示卷烟销量随云南省地区生产总值的增加而减少，从实际情况来看，2008~2018 年云南省烟草商业系统卷烟销量没有呈现持续增长的状态，说明市场需求量已基本饱和，而地区生产总值随经济发展在不断增加，卷烟所带来的经济发展在地区生产总值中的比重在不断减少。其他条件不变的情况下，云南省地区生产总值每增加 1 亿元，卷烟年销量减少 23 箱。

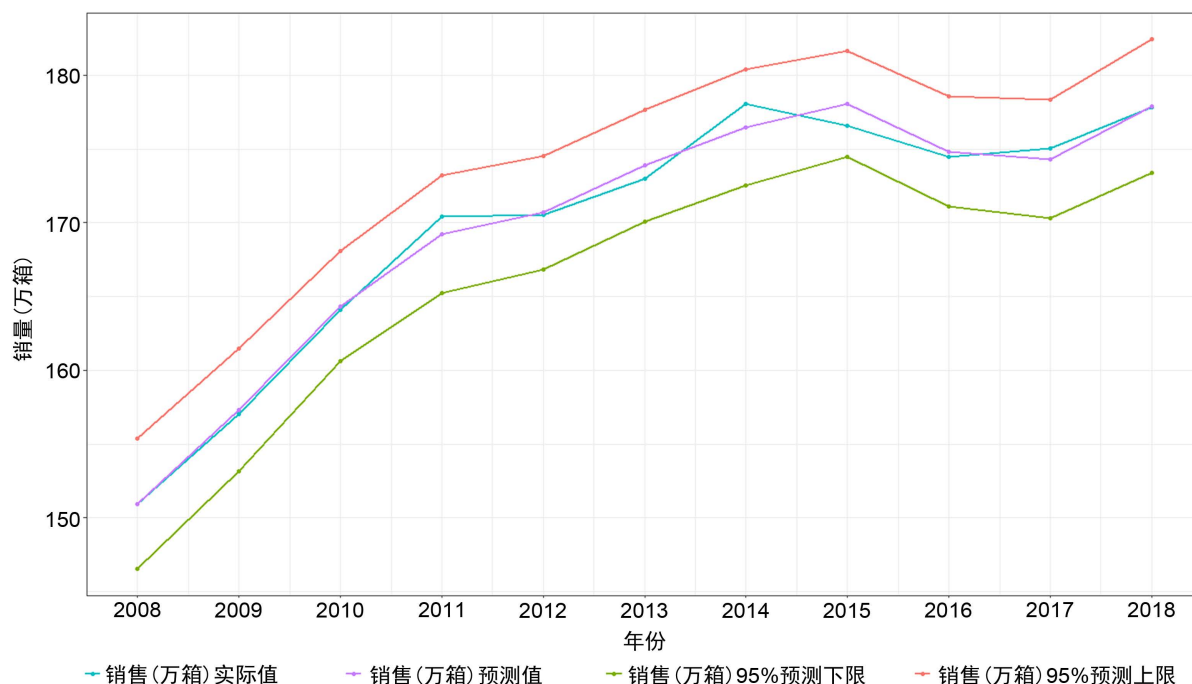


Figure 9. Curve: Fitting result

图 9. 拟合效果图

X_2 第三产业增加值占 GDP 的比重(%)、 X_3 社会消费品零售总额(亿元)、 X_5 年末常住人口(万人)的系数为正, 解释为随着第三产业增加值占 GDP 的比重、社会消费品零售总额、年末常住人口的增加, 云南省烟草商业系统卷烟销量也不断增加, 例如, 其他自变量不变的情况下, 年末常住人口每增加 1 万人, 卷烟年销量增加 1014 箱。

X_4 居民人均可支配收入(元)的系数为负, 表明一方面当人均卷烟需求量达到饱和时, 就不会再随着收入的增加而增加, 另一方面收入不断增加的同时人民的健康意识也在增强, 更多的消费选择导致卷烟购买量的减少。其他条件不变的情况下, 居民人均可支配收入每增加 1 元, 卷烟年销量减少 34 箱。

X_6 单箱销售收入(元)的系数为正, X_7 销售利税率(%)的系数为负, 可以理解为人们对价位较高的高端卷烟有更多的选择倾向, 但同时销售利税率又在抑制人们的这种消费倾向, 提高税负是有效的控烟措施。其他条件不变的情况下, 单箱销售收入每增加 1 元, 卷烟年销量增加 41 箱, 销售利税率每增加 1%, 卷烟年销量减少 3079 箱。

X_8 CPI (2008 年 = 100)的系数为负, 表示卷烟销量随着定基期 CPI 的增加而减少, CPI 代表了通货膨胀程度, 通货膨胀程度越高, 收入不变的情况下, 人们的购买力是越低的, 卷烟的销量也就随之降低。其他自变量不变的情况下, CPI (2008 年 = 100)每增加 1, 卷烟年销量就减少 7487 箱。

结合以上模型解释, 从以下三方面分析影响卷烟销量的影响因素:

第一, 社会经济因素。人口、产业结构、社会消费和通货膨胀都不同程度上影响或体现卷烟销量的变化, 随着人口、第三产业增加值占地区生产总值的比重和社会消费的增加, 卷烟销量将会随之增加, 但在卷烟市场相对饱和、产业多样化发展的实际情况下, 地区生产总值的不断增长已不再体现、带动卷烟销量的增加。而通货膨胀则通过影响购买力进而抑制卷烟销量。

第二, 人民健康因素。近年来健康、绿色、环保的理念深入人心, 人们对卷烟的危害有了更多的关注, 同时更健康的卷烟替代品市场在逐渐形成并扩张, 卷烟市场受到了一定程度的冲击。

第三，政策法规因素。政府通过政策法规对卷烟销售利税实现监督与控制，销售利税的增加进而对卷烟销量产生抑制作用，卷烟利税的提高是有效的控烟政策。

参考文献

- [1] 都恩环, 雷海潮. 提高烟草税对我国卷烟消费影响的实证研究[J]. 中国卫生经济, 2020, 39(6): 66.
- [2] 赵颖. 基于回归分析的我国汽车销量预测模型研究[D]: [硕士学位论文]. 武汉: 华中师范大学, 2014: 29-38.
- [3] 李莹. 基于时间序列与多元线性回归综合模型的农村卷烟销量预测[D]: [硕士学位论文]. 昆明: 云南大学, 2015: 33-40.