基于sas软件的我国2013年各省规模以上

工业企业资产情况分析

摘 要

本文首先将取得的我国2013年各省规模以上工业企业资产情况数据进行图形显示，初步对指标的相关性进行判断。其次基于因子分析原理在sas中对2013年中国各省（市）工业企业资产相关数据进行分析，得出了中国31个省（市）在2013年的工业企业资产发展状况综合得分。再次基于因子分析中各省得分结果对31个省（市）进行聚类，得出了31个省（市）2013年的工业企业资产发展状况分类情况。最后在sas中，使用了gmap语句尝试对此前分析出的各省发展状况结果进行地图化、三维的立体显示结果。

关键词：规模以上工业企业；sas统计地图；因子分析；聚类分析；数据可视化；

引言

工业发展状况是我国经济发展的一项重要指标，中国国家统计局每年都要对我国不同地区相关工业企业的发展情况进行调查统计，最终将结果出示在中国国家统计局统计年鉴中。如何将这些数据进行分析利用将会很大程度上体现这些统计数据的价值。

从取得的我国2013年各省规模以上工业企业资产情况数据来看，我国在2013年的工业统计数据在每个省（市）都有所差异，本文利用因子分析法从宏观上对31个省（市）2013年的工业企业资产表现进行评价排名，将我国2013年各省规模以上工业企业资产情况因子得分结果进行聚类分析，最后就因子得分结果进行了sas三维统计地图的可视化展示，使得2013年31个省（市）的工业企业资产发展状况有了直观的展示方式。这样的展示流程可以用在每一年的年度数据可视化报表中，清晰直观的展示各地区每年工业发展状态。

本文的创新之处在于采用了不同的数据展示结果，sas中的gmap语句使得区域数据的可视化变为可能，并且使得结论可以直观展现。

1. 模型方法

（一）知识准备

（1）规模以上企业

国家对不同行业的企业都制订了一个规模要求，达到规模要求的企业就称为规模以上企业，规模以上工业企业在2010年之前是指年主营业务收入在500万元及以上的[法人](http://baike.baidu.com/view/10873.htm" \t "_blank)工业企业；2011年是指年主营业务收入在2000万元及以上的法人工业企业。

（2）因子分析

因子分析法是通过对样本相关阵的内部依赖关系的研究，将一些具有一定关系的变量或样本归结为较少的几个不可观测的综合因子(又称主因子)的多元统计分析方法。首先由相应计算确定主因子的个数，建立因子模型，如果求出主因子解后，各个主因子的解释力度不够，还需要进行因子旋转，以得到一个更有解释意义的因子结构，最后应用因子分析模型去评价每年的中国规模以上企业资产总额在整个模型中的地位，即进行综合评价。

（3）聚类分析

聚类分析就是用数学方法研究和处理给定对象的分类。作为多元统计方法的一种，聚类分析又称为群分析，它是研究样品或指标的分类问题。聚类分析的研究不仅具有重要的理论意义，也具有重要的工程应用价值和人文价值。随着我们知识水平的不断提高，在实际生活和工作中，面对经常遇见分类问题，要求也越来越严格。例如，核算分析我国31个省市自治区工业企业效益等。

（4）sas统计地图

在sas软件中，除了系统自带的sasuser以及sashelp数据库中的现有数据之外，还存在一种专用于绘制统计读地图的maps数据库，数据库中存储了来自纾解各地区域图形的统计数据，其中也包含了中国完整轮廓地图，只要利用相关数据进行地图连接便可以绘制出不同种类的地图，有block map、prism map 等等，本文中使用的是prism map。

（二）因子分析

第一步：将31个省（市）2013年工业企业资产相关数据进行分析。首先对反应工业企业资产情况的企业单位数、资产总计、流动资产总计、负债合计、所有者权益合计、主营业务收入、主营业务成本、主营业务税金及附加、利润总额、本年应交增值税、总资产贡献率、资产负债率、流动资产周转次数、工业成本费用利润率一共14种指标进行初步趋势预判，在sas中绘制多线图来观察各个指标的变化趋势；

第二步：对14个指标之间的相关系数进行分析，编程运算出各个指标之间的相关系数，将绝大部分相关系数大于0.3的指标留下，将剩余指标排除；

第三步：在matlab软件中对剩余指标进行因子分析，利用matlab对应的因子分析语句进行处理，将因子得分情况输出在工作框中。

（三）聚类分析

第一步：在sas中将因子分析结果作为参照变量（指标），对31个省进行2013年规模以上工业企业资产运营状况的编程聚类分析。

第二步：针对聚类分析结果进行解释。

（四）数据可视化

第一步：将2013年全国31个省规模以上工业企业资产运营能力综合得分数据作为基础数据，并以sas软件中地图数据集为参考进行数据集的构建。

第二步：在sas中编程实现数据的prism map可视化展示。

第三步：对展示结果进行结合实际的分析。

二．结果分析

（一）指标对应说明[[1]](#footnote-1)

|  |  |
| --- | --- |
| 符号 | 指标含义 |
| place | 地区 |
| num | 企业单位数 |
| sum | 资产总计 |
| flu | 流动资产总计 |
| own | 负债合计 |
| power | 所有者权益合计 |
| prim | 主营业务收入 |
| prima | 主营业务成本 |
| tax | 主营业务税金及附加 |
| income | 利润总额 |
| xtax | 本年应交增值税 |
| rgive | 总资产贡献率 |
| rown | 资产负债率 |
| time | 流动资产周转次数 |
| rmon | 工业成本费用利润率 |

表1：表示符号对应指标查明表[[2]](#footnote-2)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 北 京 | 天 津 | 河 北 | 山 西 | 内蒙古 | 辽 宁 | 吉 林 | 黑龙江 | 上 海 | 江 苏 | 浙 江 | 安 徽 | 福 建 | 江 西 | 山 东 |

表2：数字代表区域意义信息表（聚类分析部分）

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 |
| 河 南 | 湖 北 | 湖 南 | 广 东 | 广 西 | 海 南 | 重 庆 | 四 川 | 贵 州 | 云 南 | 西 藏 | 陕 西 | 甘 肃 | 青 海 | 宁 夏 | 新 疆 |

表3：数字代表区域意义信息表（聚类分析部分）（续表）

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 安 徽 | 浙 江 | 江 西 | 江 苏 | 吉 林 | 青 海 | 福 建 | 黑龙江 | 河 南 | 河 北 | 湖 南 | 湖 北 | 新 疆 | 西 藏 | 甘 肃 | 广 西 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |

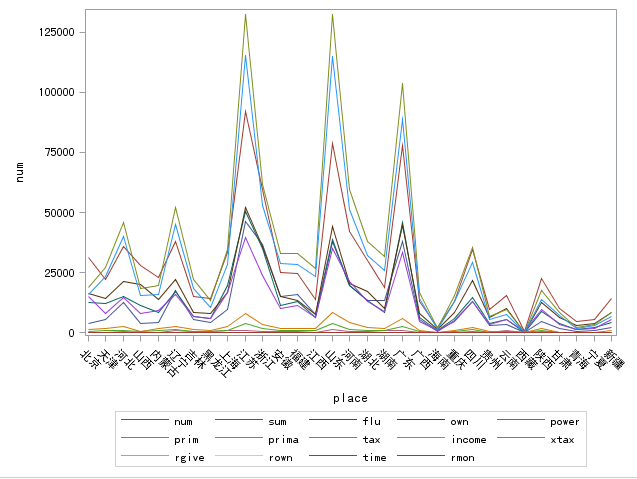
表4：数字代表区域意义信息表（数据可视化部分）

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 贵 州 | 辽 宁 | 内蒙古 | 宁 夏 | 北 京 | 上 海 | 山 西 | 山 东 | 陕 西 | 天 津 | 云 南 | 广 东 | 海 南 | 四 川 | 重 庆 |
| 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 28 | 29 | 30 | 31 | 32 | 33 |

表5：数字代表区域意义信息表（数据可视化部分）（续表）

（二）指标走势图

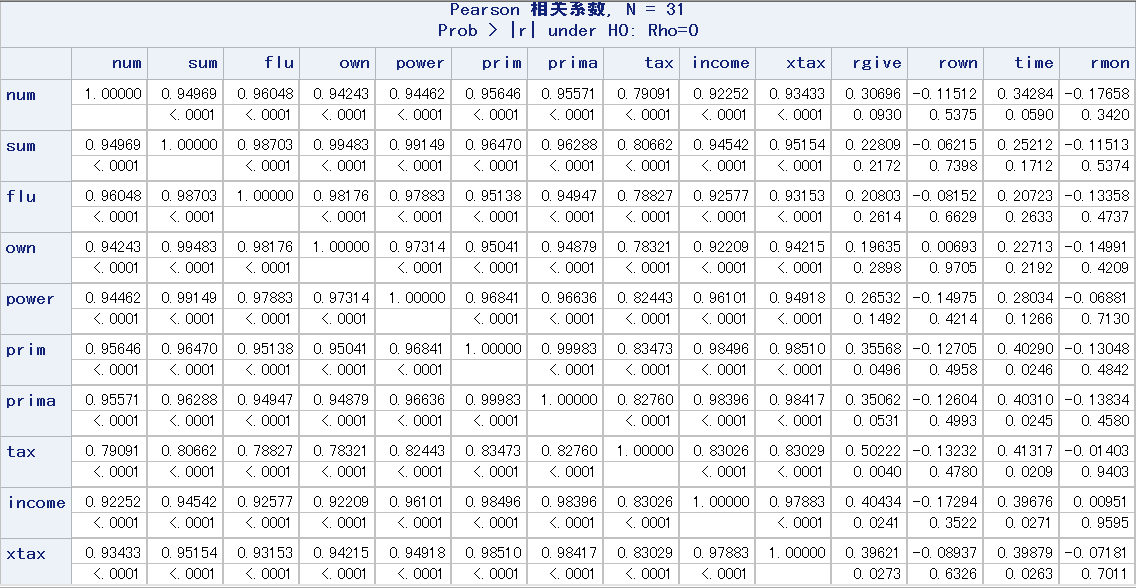
图1:2013年我国31个省（市）规模以上工业企业资产运营指标走势图



由图1我们可以发现在2013年我国31个省（市）规模以上工业企业资产运营指标中，绝大部分指标的走势都是一致的，31个省（市）在指标值方面大小差异有大有小，但是基本取向一致，没有异常数据出现。

（三）指标相关系数

图2：指标相关系数图（显著部分）



由图中数据可知，通过变量相关性检验并且效果较好的指标为：企业单位数、资产总计、流动资产总计、负债合计、所有者权益合计、主营业务收入、主营业务成本、主营业务税金及附加、利润总额、本年应交增值税。

（四）指标因子分析

图3：因子分析因子贡献率表



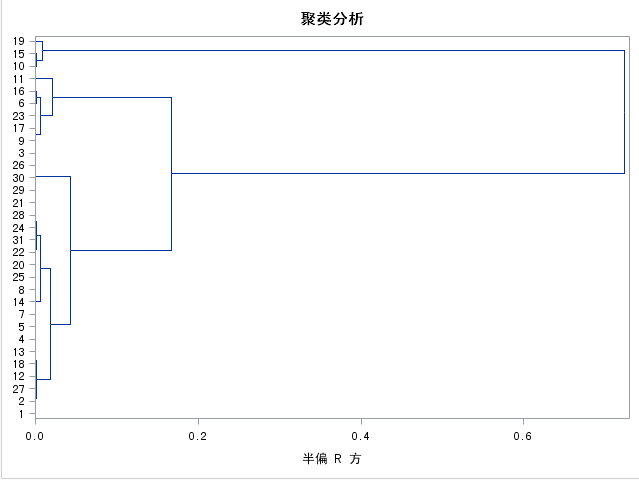
由上表可以看出，在提取第一个主因子的同时，该因子对指标的贡献率就已经达到了93.94%,这样的累计贡献率已经超过85%，它对每一个单独的剩余因子的贡献最小为85.52%，大多数集中在97%-98%之间，这样的贡献率效果说明该主成分已经可以用来作为唯一主因子进行因子分析。利用该主因子打出的全国31个省规模以上工业企业资产运营能力综合得分如下：

图4：各省（市）在2013年规模以上工业企业资产运营综合得分表



（五）指标聚类分析

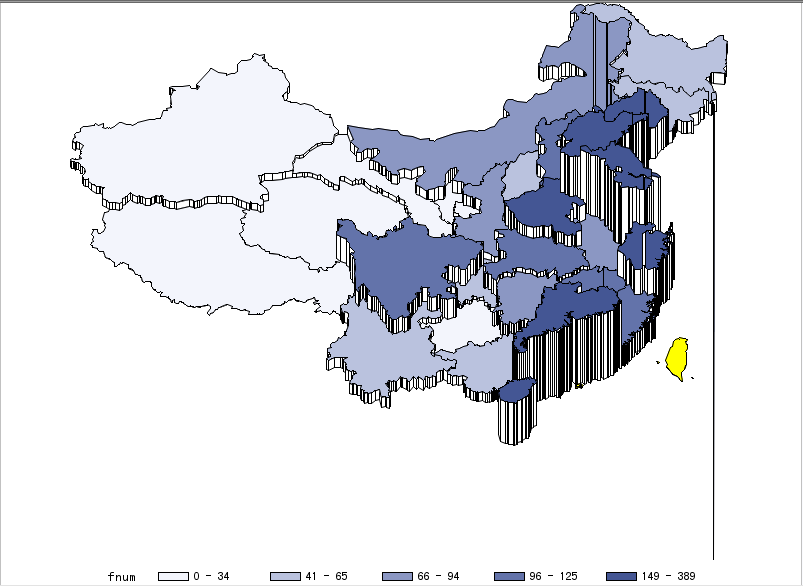
对中国2013年31个省规模以上工业企业资产运营能力综合得分的分析我们在可视化之后具体进行，接下来是由因子打分结果得出的对中国2013年31个省规模以上工业企业资产运营能力聚类分析结果：

图5:2013年中国31个省（市）规模以上企业运营能力聚类结果图

将结果进行解读，则可以看到，中国2013年31个省规模以上工业企业资产运营能力大致可以聚为三类，第一类是由江苏、山东、广东构成的规模以上工业企业运营能力较强实力队，第二类是由浙江、河南、辽宁、四川、湖北、上海构成的规模以上工业企业运营能力中实力队，剩余包括新疆、西藏、安徽、陕西在内的规模以上工业企业运营较弱实力队。

（六）分析结果可视化

图6：中国2013年31个省规模以上工业企业资产运营能力综合得分统计地图



从上图来看，在2013年，我国各个省市规模以上工业企业资产运营能力差异显著，其中，可以明显看出在沿海东部地区的省份规模以上工业企业资产运营能力相对较好，而在西部地区规模以上工业企业资产运营能力整体较弱。

按照具体观察来说，我国在2013年江苏、山东、广东省的规模企业资产运营效果明显好于其它省份，这样的观察结果与聚类分析结果相一致。而在整体实力相对较弱的西部地区，新疆省的表现算是西部地区的首位。

结合实际我们可以知道，新疆地区的规模以上工业企业主要以轻工业为主，例如在2012年，新疆的塑料节水器材产业已发展成为中国最重要的产业基地，产量世界第一，类似这样的产业为新疆的规模以上工业企业提供了发展的前景，也很大程度上拉动了新疆地区工业产业的资产运营实力。

三．结论

通过全文的分析，我们可以较为全面的得到中国2013年各个省（市）在规模以上工业企业资产运营方面具体高低、所属类别的情况，并且可以直观的感受不同省（市）2013年规模以上工业企业资产运营能力高低的区别。

在所有的31个省（市）中，一共可以将规模以上工业企业资产运营能力分为三等，其中江苏、广东、山东三个东部沿海城市居于资产运营首位，这与三个省的地理位置关系密不可分，另外，居于西部的省份大多处于资产运营能力的最底层，仍然无法与中东部省份相抗衡，但在西部省份中，新疆省凭借自身开发的轻工业规模以上企业优势超越了西部大部分省份的资产运营能力。

从研究过程看，本文的研究方法还存在一些缺陷，比如在数据可视化的阶段，将各个省（市）综合得分处理为正值的时候仅仅运用了集体加100的方式，这样使得数据间相对大小不变但相对比例发生了变化，并不是十分科学的展现省（市）之间的资产运营能力比例情况。

但是本文比较好的一点就是尝试了使用sas丰富的数据库资源，并且采用了可视化的方式来展示数据分析结果，使得数据分析结论不再是一堆数据，能够帮助我们更好的直观观察数据分析结果。

附录

1. **sas相关代码数据输入部分**

libname china "d:\baogao";

data china.map;

input place $ num sum flu own power prim prima tax income xtax rgive rown time rmon;

cards;

北京 3701 31398.28 12772.76 16363.61 15034.67 18624.82 15798.25 282.41 1254.78 521.77 7.66 52.12 1.54 7.09

天津 5383 22059.41 12159.96 14095.77 7963.64 27011.12 23208.38 317.95 1992.76 859.33 16.09 63.90 2.36 8.02

河北 12649 36040.17 14970.43 21164.61 14875.56 45766.25 40019.94 423.24 2560.86 1112.99 13.56 58.73 3.20 6.02

山西 3946 28058.27 11371.37 20011.00 8047.27 18404.65 15687.34 162.61 547.91 735.30 7.33 71.32 1.65 3.09

内蒙古 4377 23141.71 8241.90 13893.68 9248.03 19550.83 15732.66 272.56 1682.55 792.40 13.92 60.04 2.41 9.73

辽宁 17561 37989.29 17215.31 22219.96 15769.33 52150.40 45093.21 856.54 2461.58 1250.00 14.14 58.49 3.22 5.06

吉林 5353 15257.90 6676.97 8354.55 6903.35 21950.72 18396.65 496.10 1230.10 610.06 17.47 54.76 3.51 6.06

黑龙江 4098 14059.17 5879.75 8033.28 6025.89 13569.81 10694.46 665.03 1150.21 603.47 19.08 57.14 2.38 9.78

上海 9782 33538.26 19698.20 16885.77 16652.49 34533.53 28569.00 910.66 2415.20 905.28 13.71 50.35 1.82 7.59

江苏 46387 92081.69 50358.02 52286.71 39794.98 132270.41 115111.59 1127.44 7834.06 3985.16 16.22 56.78 2.75 6.34

浙江 36904 59633.11 34845.62 35787.46 23845.65 61765.48 53100.85 671.26 3385.87 1627.38 11.83 60.01 1.85 5.82

安徽 15114 25168.07 11281.46 14957.10 10210.97 33079.46 28607.39 420.87 1758.77 866.98 14.46 59.43 3.15 5.69

福建 15806 24671.06 12960.93 13489.28 11181.78 32847.14 28241.09 376.94 1959.45 916.88 15.97 54.68 2.76 6.41

江西 7601 13640.12 6233.24 7402.14 6237.98 26700.22 23246.90 272.33 1756.66 853.41 24.38 54.27 4.60 7.19

山东 38654 78881.06 37951.36 44011.52 34869.54 132318.98 114856.58 1483.74 8507.73 3698.52 20.21 55.79 3.74 6.97

河南 19773 42021.92 19504.21 20506.42 21515.50 59454.79 51193.95 640.54 4410.82 1436.60 18.49 48.80 3.38 8.13

湖北 13441 30131.82 13595.46 16968.39 13163.43 37864.54 32146.59 733.42 2080.66 946.77 15.00 56.31 3.03 5.91

湖南 13323 19031.64 8326.78 10284.76 8746.88 31616.57 25830.03 849.25 1585.06 1123.73 21.64 54.04 4.11 5.62

广东 38094 77943.52 45844.47 44656.62 33286.90 103654.98 88559.82 1144.61 5854.93 2539.07 13.85 57.29 2.42 6.03

广西 5396 13063.37 6152.22 8215.16 4848.21 16726.00 14228.60 318.67 874.00 528.11 15.70 62.89 2.96 5.65

海南 391 2328.02 948.98 1243.36 1084.66 1640.70 1312.75 80.92 110.80 62.78 12.89 53.41 1.87 7.60

重庆 5237 13135.92 6127.30 8315.09 4820.83 15417.07 13149.80 225.09 878.43 579.91 15.57 63.30 2.71 6.10

四川 13163 34729.16 14688.47 21804.19 12924.97 35251.84 29353.20 575.96 2168.37 1244.38 14.10 62.78 2.52 6.71

贵州 3139 9703.64 3900.19 6155.22 3548.42 6878.40 5383.61 286.55 477.33 313.83 13.96 63.43 1.93 7.77

云南 3382 15344.41 5717.80 9918.34 5426.07 9773.14 7587.10 786.69 549.08 440.13 14.59 64.64 1.79 6.43

西藏 70 548.63 138.56 186.68 361.95 93.37 80.15 1.54 7.17 8.23 3.66 34.03 0.67 7.48

陕西 4489 22443.11 8775.97 12581.24 9861.87 17763.00 13753.67 522.72 1973.32 856.04 16.97 56.06 2.04 12.86

甘肃 1830 10159.43 3959.31 6537.52 3621.91 8443.65 7320.96 266.38 286.71 192.28 9.54 64.35 2.22 3.63

青海 465 4597.68 1414.56 3045.69 1551.99 2045.38 1632.58 47.40 141.34 90.91 8.23 66.24 1.53 7.58

宁夏 935 5588.03 2049.39 3717.57 1870.46 3374.49 2879.49 68.41 139.11 99.65 8.06 66.53 1.77 4.37

新疆 2102 14238.01 4462.98 8615.65 5622.36 8608.03 6745.84 329.83 795.40 329.47 12.82 60.51 2.13 10.64

;

proc print;

run;

1. **sas相关代码数据走势图部分**

proc sgplot data=china.map;

series x=place y=num;

series x=place y=sum;

series x=place y=flu;

series x=place y=own;

series x=place y=power;

series x=place y=prim;

series x=place y=prima;

series x=place y=tax;

series x=place y=income;

series x=place y=xtax;

series x=place y=rgive;

series x=place y=rown;

series x=place y=time;

series x=place y=rmon;

run;

/\* or\*/

data map;

set china.map;

array arrayr[\*] ra--rd;

do i=1 to dim(arrayr);

r=arrayr[i];

group=i;

output;

end;

run;

proc sgplot data=map;

series x=place y=r/group=group;

run;

1. **sas相关代码因子分析部分**

proc corr data=china.map cov;

var num sum flu own power prim prima tax income xtax rgive rown time rmon;

run;

PROC FACTOR DATA=china.map n=1 out=china.outf;

var num sum flu own power prim prima tax income xtax;

RUN;

data china.fscore;

set china.outf;

keep place factor1;

proc print;

run;

1. **sas相关代码聚类分析部分**

proc cluster data=china.fscore

method=ward;

var factor1;

proc tree;

run;

1. **sas相关代码数据可视化准备部分**

data china.final;

input id fnum density;

datalines;

22 76 1

28 77 2

10 125 3

24 65 4

20 66 5

19 149 6

5 58 1

8 53 3

23 125 1

4 389 2

2 219 3

1 94 4

7 96 5

3 96 6

25 363 2

9 170 3

12 115 2

11 93 1

30 315 4

16 44 3

31 204 6

33 41 6

32 119 1

18 24 4

29 50 1

14 0 3

26 73 4

15 22 5

6 0 3

21 5 4

13 34 5

;

proc print;

run;

1. **sas相关代码数据可视化执行部分部分**

goptions reset=all border;

\*title1 "Population in Africa";

\*footnote1 j=r "GMPPRISM";

proc gmap data=china.final

map=maps.china density=low all;

id id;

prism fnum / cdefault=yellow relzero;

run;

quit;

1. 由于部分指标长度过长，显示在结果输出界面时使得图表过于臃肿，截图遇到困难，所以这里将指标和地名统一加以符号化 [↑](#footnote-ref-1)
2. 由于在《2013年中国统计年鉴》中各省市排名的次序与sas软件中自带的maps数据库各省数据顺序不同，所以这里单独列出相关关系以免混淆 [↑](#footnote-ref-2)