

成人收入影响因素分析 (Lasso)

本文使用 Lasso 方法，使用 cgss2015 年数据，和 lasso 分析方法，对成人收入影响因素进行筛选。总共选择了 104 个变量，具体变量将附在文章最后。分别使用了 lasso 模型中使用 cross-validation 中自定义的惩罚程度、0.1 惩罚度、0.2 惩罚度。

当为程序使用 cross-validation 中惩罚程度为 0.0582 时，筛选变量得到 31 个；使用惩罚度为 0.1 时，筛选变量得到 21 个；使用惩罚度为 0.1 时，筛选变量得到 8 个。其中最重要的变量为：性别、教育程度、户口、业余学习、与邻居社交时间、工作性质等。

使用的 stata 命令：

```
elasticregress lna8a    ///
s41  a2 a301 a4 a501 a7a a7b a10 a11 ///
a121  a15 a13 a14 a16 a17 a18 a29 a3001 a3012 /// *a19
a311 a312 a313 a31a a31b a32 a33 a34 a35 a36 a371 a38 a39 a40 a421 ///
a424 a431 a432 a433 a434 a44 a45 a46 a48 a51 a50 a52 /// *a5699 a5601 a53 a58    ///
a611 a612 a613 a614 a63 a64 a65    ///
a66 a69 a71a a72 a73 a75a a75b a76    a89b a89c a89d a89f a89g ///
b1005 b1006 b1007 b1008 b1009    b1013 b111 b112 b12 b131 b132 ///
b134 b141 b142 b143 b161 b162 b163 b164 ///
fisco88 misco88, alpha(1)
```

另外后面两个惩罚度使用：alpha(1) lamdba(0.1)；alpha(1) lamdba(0.2)。

输出命令使用：eststa 命令。

具体结果如下表：

Table1:elasticregress:lasso			
	CV	lamdba(0.1)	lamdba(0.2)
s41	0	0	0
a2	-1.110	-1.080	-1.004
a301	0	0	0
a4	0.00977	0	0
a501	0.141	0	0
a7a	0.0477	0.0473	0.0416
a7b	0	0	0
a10	0	0	0
a11	-0.00108	0	0
a121	0.152	0.0881	0
a15	0.0720	0.0285	0
a13	0	0	0
a14	0	0	0
a16	0.0384	0.0437	0
a17	0.0354	0.00206	0

a18	0.182	0.175	0.151
a29	0	0	0
a3001	0.0444	0.0115	0
a3012	0	0	0
a311	0	0	0
a312	-0.0347	0	0
a313	0.0801	0.0727	0.0447
a31a	0.0546	0.0451	0.00234
a31b	0	0	0
a32	0.0708	0.0378	0
a33	0	0	0
a34	0	0	0
a35	0	0	0
a36	0	0	0
a371	-0.00570	0	0

a38	0	0	0
a39	0	0	0
a40	0	0	0
a421	0	0	0
a424	0	0	0
a431	0.0484	0.0330	0
a432	0	0	0
a433	0	0	0
a434	0	0	0
a44	0	0	0
a45	-0.210	-0.200	-0.157
a46	0	0	0
a48	0.0197	0	0
a51	0	0	0
a50	0	0	0
a52	0	0	0

a58	-0.789	-0.762	-0.696
a611	0	0	0
a612	-0.114	-0.0620	0
a613	0	0	0
a614	0	0	0
a63	0	0	0
a64	0.468	0.459	0.415
a65	0	0	0
a66	0	0	0
a69	0	0	0
a71a	-0.0127	-0.00787	0
a72	0	0	0
a73	0	0	0
a75a	0	0	0
a75b	0	0	0

a76	-0.0941	-0.0720	0
a89b	0	0	0
a89c	0	0	0
a89d	0	0	0
a89f	0	0	0
a89g	0	0	0
a90b	0	0	0
a90c	0	0	0
a90d	0	0	0
a90f	0	0	0
a90g	0	0	0
b1	0	0	0
b2	-0.142	-0.0869	0
b3	-0.0479	0	0
b4a	-0.0163	0	0

b5	0	0	0
b8a	0.00000782	0	0
b8b	0	0	0
b1001	-0.0261	0	0
b1002	0	0	0
b1003	0	0	0
b1004	0	0	0
b1005	0	0	0
b1006	0	0	0
b1007	0.0833	0	0
b1008	0	0	0
b1009	0	0	0
b1013	0	0	0
b111	0	0	0
b112	0	0	0
b12	0	0	0

b131	0	0	0
b132	0	0	0
b134	-0.0499	-0.00405	0
b141	0	0	0
b142	0	0	0
b143	0	0	0
b161	0	0	0
b162	0	0	0
b163	0	0	0
b164	0	0	0
fisco88	0	0	0
misco88	0	0	0
_cons	35.74	26.14	10.78
<i>N</i>	2134	2134	2134
<i>R</i> ²	0.316	0.300	0.277
lambda	0.0582	0.100	0.200

t statistics in parentheses

Source: cgss2015_12_modified.dta

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

变量与标签

- a2 "您的性别"
- a301 "您的出生日期-年"
- a4 "您的民族"
- a501 "您的宗教信仰-不信仰宗教"
- a7a "您目前的最高教育程度（包括目前正在读的）"
- a7b "您目前的最高教育程度的状态"
- a10 "您目前的政治面貌"
- a11 "您现在住的这座住房的套内建筑面积"
- a121 "您现在这座房子的产权属于-自己所有"
- a15 "您觉得您目前的身体健康状况"
- a13 "您目前的身高是（厘米）"
- a14 "您目前的体重是（斤）"
- a16 "在过去的四周中由于健康问题影响到您的工作或其他日常活动的频繁程度"
- a17 "在过去的四周中您感到心情抑郁或沮丧的频繁程度"
- a18 "您目前的户口登记状况"
- a29 "以上媒体中哪个是您最主要的信息来源"
- a3001 "过去一年，您是否常在空闲时间从事以下活动-看电视或看碟"
- a3012 "过去一年，您是否常在空闲时间从事以下活动-上网"
- a311 "在过去一年中，您是否经常在您的空闲时间做下面的事情-社交"
- a312 "在过去一年中，您是否经常在您的空闲时间做下面的事情-休息放松"
- a313 "在过去一年中，您是否经常在您的空闲时间做下面的事情-学习充电"
- a31a "您与邻居进行社交娱乐活动的频繁程度"
- a31b "您与其他朋友进行社交娱乐活动的频繁程度"
- a32 "过去一年，您有多少个晚上是因为出去度假或者探访亲友而没有在家过夜"
- a33 "总的来说，您同不同意在这个社会上，绝大多数人都是可以信任的"
- a34 "总的来说，您同不同意在这个社会上，您一不小心，别人就会想办法占您的便宜"
- a35 "总的来说，您认为当今社会公不公平"
- a36 "总的来说，您觉得您的生活是否幸福"
- a371 "如果没有政策限制的话，您希望有几个孩子"
- a38 "您认为婚前性行为对不对"
- a39 "您认为婚外性行为对不对"
- a40 "您认为同性间的性行为对不对"
- a421 "您是否同意-男人以事业为重，女人以家庭为重"
- a424 "您是否同意-在经济不景气时，应该先解雇女性员工"
- a431 "您认为自己目前处于哪个等级上"
- a432 "您认为您10年前处于哪个等级上"
- a433 "您认为您10年后将会在哪个等级上"

a434 "您认为在您14岁时，您的家庭处在哪个等级上"

a44 "上次居委会选举/村委会选举，您是否参加了投票 "

a45 "您是不是工会会员"

a46 "您是否同意以下说法-如果有人在公共场所发布批评政府的言论，政府不应该干涉"

a48 "您是否同意以下说法-在哪里工作和生活是个人的自由，政府不应该干涉"

a51 "您觉得自己听英语的能力是什么水平"

a50 "您觉得自己说普通话的能力是什么水平"

a52 "您觉得自己说英语的能力是什么水平"

a58 "您的工作经历及状况"

a611 "您目前是否参加了以下社会保障项目-城市基本医疗保险/新型农村合作医疗保险/公费医疗"

a612 "您目前是否参加了以下社会保障项目-城市/农村基本养老保险"

a613 "您目前是否参加了以下社会保障项目-商业性医疗保险"

a614 "您目前是否参加了以下社会保障项目-商业性养老保险"

a63 "您家目前住在一起的通常有几人（包括您本人）"

a64 "您家的家庭经济状况在所在地属于哪一档"

a65 "您家现拥有几处房产"

a66 "您家是否拥有家用小汽车"

a69 "您目前的婚姻状况"

a71a "您目前的配偶或同居伴侣是哪一年出生的"

a72 "您配偶或同居伴侣目前的最高教育程度"

a73 "您配偶或同居伴侣的政治面貌"

a75a "您配偶或同居伴侣去年全年的总收入"

a75b "您配偶或同居伴侣去年全年的职业/劳动收入"

a76 "您配偶或同居伴侣上一周的就业状况"

a89b "您父亲的最高教育程度"

a89c "您父亲的政治面貌 "

a89d "您14岁时父亲的就业状况"

a89f "您14岁时父亲的职务级别"

a89g "您14岁时父亲工作的单位或公司的类型"

a90b "您母亲的最高教育程度"

a90c "您母亲的政治面貌 "

a90d "您14岁时母亲的就业状况"

a90f "您14岁时母亲的职务级别"

a90g "您14岁时母亲工作的单位或公司的类型"

b1 "与同龄人相比，您本人的社会经济地位怎样"

b2 "与三年前相比，您的社会经济地位发生了什么变化"

b3 "根据您的印象您对一些重要事情所持的观点和看法与社会大众一致的时候"

有多少"

b4a "以下工作回报您最希望获得"

b5 "考虑到您的能力和工作状况，您目前的收入是否合理"

b8a "您认为每月户平均收入低于多少元就属于困难户了"

b8b "您认为每月户平均收入高于多少元就属于富裕户了"

b1001 "在不直接涉及金钱利益的一般社会交往/接触中的信任度-(近)邻居"

b1002 "在不直接涉及金钱利益的一般社会交往/接触中的信任度-(城镇的)远邻/街坊或(乡村)邻居?"

b1003 "在不直接涉及金钱利益的一般社会交往/接触中的信任度-同村的同姓人士"

b1004 "在不直接涉及金钱利益的一般社会交往/接触中的信任度-同村的非同姓人士"

b1005 "在不直接涉及金钱利益的一般社会交往/接触中的信任度-亲戚"

b1006 "在不直接涉及金钱利益的一般社会交往/接触中的信任度-同事"

b1007 "在不直接涉及金钱利益的一般社会交往/接触中的信任度-交情不深的朋友/相识"

b1008 "在不直接涉及金钱利益的一般社会交往/接触中的信任度-老同学"

b1009 "在不直接涉及金钱利益的一般社会交往/接触中的信任度-在外地相遇的同乡(以同市或同县?"

b1013 "在不直接涉及金钱利益的一般社会交往/接触中的信任度-陌生人"

b111 "您对以下看法同意程度-"未婚同居"是个人行为，他人不应该指责"

b112 "您对以下看法同意程度-"同性恋"是个人行为，他人不应该指责"

b12 "您和邻居，街坊/同村其他居民互相之间的熟悉程度"

b131 "您是否同意以下说法-应该从有钱人那里征收更多的税来帮助穷人"

b132 "您是否同意以下说法-现在有的人挣的钱多，有的人挣的少，但这是公平的"

b134 "您是否同意以下说法-在我们这个社会，工人和农民的后代与其他人的后代一样，有同样多?"

b141 "总体来说，您认为政府是否应该或有责任提供以下福利-人人有工作机会"

b142 "总体来说，您认为政府是否应该或有责任提供以下福利-人人有医疗保险"

b143 "总体来说，您认为政府是否应该或有责任提供以下福利-为老人提供生活保障"

b161 "综合考虑各个方面，您对于公共教育服务的总体满意度"

b162 "综合考虑各个方面，您对于医疗卫生公共服务的总体满意度"

b163 "综合考虑各个方面，您对于基本住房保障公共服务的总体满意度"

b164 "综合考虑各个方面，您对于社会管理公共服务的总体满意度"

fisco88 "父亲职业"

misco88 "母亲职业"