

# 领导人个人特质与政府决策的相关性

自 64 郭旭东 2016011451

## 摘要:

本文运用概率论与数理统计的知识,借助统计软件 R,对领导人个人特质(性别、受教育程度、年龄)与政府决策(公共品供给水平,以教育在财政支出中所占比例衡量)间的相关性进行了分析,采用了多元线性回归、偏相关系数分析、方差分析等手段,一步步优化模型,最终得出二者之间的确存在相关性的结论。

## 关键词:

多元线性回归、偏相关系数、方差分析、执政风格、公共品供给

## 一、问题背景

地方政府在辖区的经济发展、居民福利等执政目标之间会分配不同的权重,以制定合适的财政政策。与此同时,地区领导人的个人特征,如性别、教育背景、年龄等因素可能会影响其权重分配。这两者之间是否存在某种关联性呢?是否意味着不同性别、不同学历、不同阅历的政治家会形成不一样的执政风格呢?

经查阅资料,国内外对这个问题已经有了一定研究。以性别为例,大量研究认为女性领导人具有一定的政策偏好,甚至一些国家为了提高政府性别平等和政策的多元化,颁布性别配额政策,规定了某一职位中女性的最低比例。比如说,瑞典于 1994 年颁布一项政策,要求基层政府中女性的比例不得低于 50%。

在此我希望运用概率论与数理统计的知识,对这个问题进一步深入研究。

## 二、建立回归模型

我们近似认为地区政府财政支出主要由经济建设支出与公共品供给支出构成,公共品供给代表着社会福利在政府决策模型中所占的比重。其中,教育是地方政府提供的最重要的公共品之一,于是可以用财政支出中教育支出所占的比重衡量公共品供给水平。

在领导人个人特质方面,以中国地级市市委书记作为样本,选取性别、年龄、受教育程度作为领导人特质进行研究。其中,市委书记若为女性,则性别这一项设定为 1,男性为 0。受教育程度从中专、高中、大专、本科、硕士、博士、博士后依次被赋值为 1~7,也就是说,受教育程度越高的市委书记该变量的值越大。

对于领导人特质和公共品之间的实证检验,只有在控制了一系列相关变量之后才能得到。这些变量包括人均 GDP、非农业人口比例等。

经查阅资料,在类似问题的研究中,一般采用线性模型对各变量进行近似拟合。

综上,可以得到所需要的多元线性回归模型:

$$Y = \beta_0 + \beta_1 C_1 + \beta_2 C_2 + \beta_3 X_1 + \beta_4 X_2 + \beta_5 X_3 + \varepsilon$$

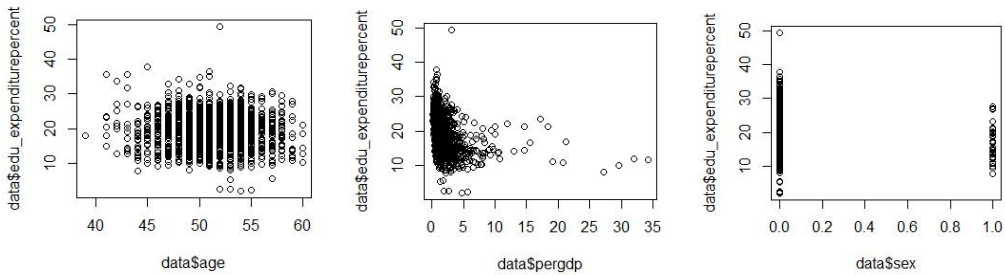
其中  $Y$  为因变量，即财政支出中教育支出所占的比重， $C_k$  为控制变量，即人均 GDP、非农业人口比例， $X_i$  为自变量，即地区领导人的性别、年龄、受教育程度。 $\varepsilon$  为随机误差，满足正态分布。

### 三、数据的获取与清洗

本文的数据来源为：教育支出占财政支出比例、人均 GDP、非农业人口比例来源于国泰安数据库(<http://www.gtarsc.com/Home>)。市委书记数据来自《Data on Prefectural Party Secretary and Mayor of P. R. China: 2000-2010》。样本的时间跨度为 2000 至 2010 年，覆盖全国全部地级市（不含北京、上海、天津及重庆四个直辖市）。

在对原始数据中的异常值、缺省项进行去除后，得到清洗后数据，使用 R 进行 outlier 检验，可以保证其中没有离群值。清洗后样本总数为 1402，足够大到分析出可靠的结论。

### 三、回归分析



由以上因变量与几个控制变量或自变量的散点图可以初步判断，因变量与这些量中的每一个并没有非常强的单一的相关性，因而使用多变量解释的回归模型有一定可靠性。用 R 进行回归分析可知：

	系数估计值	标准差	P 值	显著程度
$b_0$	24.225939	2.509894	$<2e-16$	***
非农业人口比例	-0.127066	0.007775	$<2e-16$	***
人均 GDP	-0.094581	0.058221	0.1045	
领导人性别	-1.718231	0.739354	0.0203	*
领导人教育程度	0.055017	0.182184	0.7627	
领导人年龄	-0.011605	0.038981	0.7660	

注：显著程度： 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

可以发现，非农业人口比例与因变量有着极强的相关性，领导人的性别对因变量的影响也很大，而其他三个量在本模型中不显著，所以需要进行进一步的模型修正。

通过逐步回归分析，在每引入一个解释变量后都进行 F 检验，并对已经选入的解释变量逐个进行 t 检验，当原来引入的解释变量由于后面解释变量的引入变

得不再显著时，则将其删除，从而可以保证每个解释变量都具有足够的显著性。

由此得出的修正后的回归模型去除了领导人的年龄与受教育程度两个变量，再次进行回归分析可得：

	系数估计值	标准差	P 值	显著程度
$b_0$	23.903945	0.251593	$<2e-16$	***
非农业人口比例	-0.127194	0.007767	$<2e-16$	***
人均 GDP	-0.096372	0.057776	0.0955	.
领导人性别	-1.706899	0.738631	0.0210	*

注：显著程度：0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

各变量的显著程度均比较高。这也说明了领导人特质中对公共品供给有最大影响的因素是性别，而且影响程度足够大。

#### 四、相关分析

在此基础上，可以使用偏相关系数消去人均 GDP 以及非农业人口比例的影响，得到领导人性别与公共品供给水平之间关系更准确的刻画。

计算可得：

$$r_{14 \cdot (23)} = -0.06168774, \text{ 相应 P 值为 } 0.02098353$$

（注：下标 1234 分别代表财政支出中教育支出所占的比重、非农业人口比例、人均 GDP、性别、受教育程度）

说明在水平  $\alpha=0.05$  时，去除其他变量影响后，领导人性别对公共品供给水平的影响依然显著，而且呈现负相关，即女性领导人管理的政府对于公共品供给的投入相对更少。

对比不消去人均 GDP 以及非农业人口比例的影响而直接计算的相关系数：

$$r_{14} = -0.09032055, \text{ 相应 P 值为 } 0.0007097$$

达到了非常显著的程度，实际上这其中控制变量的影响。这也在一定程度上表明领导人性别与地区人均 GDP 以及非农业人口比例间存在着相互影响。通过计算相关系数也可以说明这一点，如：

$$r_{34} = 0.07508201, \text{ 相应 P 值为 } 0.004912$$

#### 五、方差分析

通过以上分析已经可以确定领导人性别确实会对地区公共品供给产生显著影响，进一步重新考虑年龄和受教育程度的影响。实际上，由于这两个变量是多分类变量（分类类别大于 2），用普通的回归分析可能得到不准确的结果，这也是为什么逐步回归之后得到更优的回归模型时会剔除这两个变量。改用单因素方差分析得到结果如下：

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
非农业人口比例	1	8964	8964	457.849	$<2e-16$	***
人均 GDP	1	63	63	3.195	0.0741	.
领导人性别	1	107	107	5.464	0.0196	*

领导人教育程度	5	686	137	7.006	1.78e-06	***
领导人年龄	20	436	22	1.113	0.3283	

注：显著程度：0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

与之前回归分析结果相反，受教育程度在方差分析中表现出了极强的显著性，甚至超过了性别的影响，而年龄依旧与因变量的关系不明显。剔除年龄变量后再进行方差分析：

	Df	SumSq	MeanSq	Fvalue	Pr(>F)	
领导人性别	1	303	303	15.449	8.89e-05	***
领导人教育程度	5	1076	215	10.970	2.24e-10	***
非农业人口比例	1	8386	8386	427.607	<2e-16	***
人均 GDP	1	55	55	2.821	0.0932	.

注：显著程度：0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

此时，领导人性别和受教育程度表现出的显著性进一步增强。

## 六、结论

综上所述，虽然人与人的差异是多方面的，无法对每一个特征都进行类似分析，但至少可以肯定，领导人的性别和受教育程度对因变量有显著影响，而年龄影响不大。可以得出结论，地区领导人特质确实会对以教育支出为代表的公共品供给水平产生影响，进而可以合理推论，领导人特质会形成一定的执政风格，影响到地区政府的决策和资源分配。

## 七、参考文献

- 1、陈希孺，概率论与数理统计，中国科学技术大学出版社，2009 年第 1 版
- 2、宋冉和陈广汉，官员特征、经历与地方政府教育支出偏好，经济管理，2016 年第 12 期
- 3、R 之相关性的显著性检验 <https://blog.csdn.net/u012429555/article/details/78905585>
- 4、陈硕，分税制改革、地方财政自主权与公共品供给，经济学（季刊），2010 年第 4 期。

## 八、附录

R 代码：

```
data<-read.csv("C:/Users/traceur/Desktop/data.csv",header=T)
head(data)
```

```
#散点图分析
pairs(data)
cor(data)
```

```

cor.test(data$edu_expenditurepercent, data$nonagr_per)

#回归分析
fit0<-lm(edu_expenditurepercent~., data=data)
summary(fit0)
anova(fit0)
drop1(fit0)
step(fit0)

#outlier
h<-hatvalues(fit0)
hstan<-1
count<-0
for(i in h){
  if(i>hstan)
    count<-count+1
}
count
#数据完整度高, 没有 outlier

#人均 gdp 对模型影响不显著, 去掉进行分析
fit1<-lm(edu_expenditurepercent~nonagr_per+ sex + edu + age, data=data)
summary(fit1)

#全方差分析模型
data$edu<-factor(data$edu)
data$age<-factor(data$age)
data$sex<-factor(data$sex)
fita1<-aov(edu_expenditurepercent~., data=data)
summary(fita1)
fita2<-aov(edu_expenditurepercent~sex+edu + nonagr_per+pergdp, data=data)
summary(fita2)

library(lsmmeans)
sex.means<-lsmmeans(ref.grid(fita2), "sex")
edu.means<-lsmmeans(ref.grid(fita2), "edu")
age.means<-lsmmeans(ref.grid(fita2), "age")
age.means

#偏相关系数分析
library(ggm)
library(ggm)
library(igraph)

```

```
s<-cov(data)
r<-pcor(c(1,4,2),s)
r
pcor.test(r,1,dim(data))
r<-pcor(c(1,5,2),s)
r
pcor.test(r,1,dim(data))
r<-pcor(c(4,2,1,3),s)
r
pcor.test(r,2,dim(data))
```