

期中测试考情分析 及重点讲解与练习

助教课

2025.5.12

你觉得期中测试难度如何?

- ☐ A So easy! ^_^
- ☐ B It' s okay.
- ☐ C A little bit difficult.
- ☐ D It' s too hard for me. T~T

提交

考情分析

题号	1.1	1.2	1.3	1.4	1.5	2.1	2.2	2.3	2.4	2.5
满分	5	5	5	5	5	5	5	5	5	5
均分	3.44	3.31	2.69	1.88	1.38	4.25	4.44	4.69	4.06	3.19

题号	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	3.10	4
满分	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	25
均分	2.31	2.31	2.34	2.34	2.19	2.34	1.81	1.47	1.47	0.56	2.50

2.0 数据导入

- 什么时候用bcuse, 怎样导入教材数据？
- 什么时候用use/sysuse？

- 读取ASCII数据：

*- 《基本无害的计量经济学》 chap7 分位数回归

insheet using <https://economics.mit.edu/sites/default/files/publications/jtpa.raw>, tab clear

*- 第5.2题

insheet using "student-mat.csv", delimiter(";") clear

表 2.9 insheet options 内容表

[no] double	设置变量的存储类型
tab	设置制表符为数值变量的分隔符
comma	设置逗号为数值变量的分隔符
delimiter ("char")	设置自定义的“char”为数值变量的分隔符
clear	清空内存
+ [no] names	设置变量的名称保留在文件中的第一行

上次介绍了 `use` 命令，用于打开 `.dta` 型的数据文件。这里我们再介绍两个命令用于打开、获取 `.dta` 型的数据文件，分别为 `sysuse` 和 `webuse`。前者用于从 Stata 的内置数据库（具体见表 1）中调取数据，后者用于从网络上调取数据。

表 1: Stata 内置的常用数据文件 (`.dta` 格式)

文件名（后缀省略）	数据内容	文件名（后缀省略）	数据内容
<code>auto</code>	美国 1978 年汽车数据	<code>bplong</code>	血压数据
<code>cancer</code>	药物实验生存数据	<code>census1980</code>	年美国分州人口普查数据
<code>citytemp</code>	美国城市气温数据	<code>educ99gdp</code>	教育与 GDP 关系数据
<code>gnp96</code>	美国 1967-2002 年的 GNP 数据	<code>lifeexp</code>	预期寿命数据
<code>nlsw88</code>	美国年轻妇女研究数据	<code>pop2000</code>	美国 2000 年人口普查数据
<code>sp500</code>	s&p500 公司历史数据	<code>uslifeexp</code>	美国预期寿命 1900-1999
<code>voter</code>	美国 1992 年选举民意调查数据		

注：`auto.dta` 数据库为美国 1978 年汽车数据，包括产地、车名、行使里程、重量等变量。由于它同时备份在根目录下，所以同时可以用 `use` 命令和 `sysuse` 命令打开。

而 `webuse` 命令默认从 `http://www.stata-press.com/data` 这一路径获取数据。欲获取其他网络数据则只能使用 `use + 完整的网络地址` 来获取。

1.1 变量标签

- label可以给数据集、变量、变量的值贴标签。

*- 删除原有的标签

label drop gender

*- 定义标签

label define gender 1 "男" 0 "女"

*- 给gender的值加标签

label values gender "gender"

*- 给weight变量加标签，方便查看

label variable weight "体重"

- 练习：使用auto数据定义foreign的值标签



1.2 数据类型转换

```
. deststring Accper , replace ignore("-")
Accper: character - removed; replaced as long
```

表 2.3 deststring 选项表

ignore ("chars")	删除字符型变量中的非数值型字符，如空格、特殊符号等
force	将非数值型字符转化为缺失值
float	定义新变量的类型为 float
percent	将百分数转化为分数

	Stkcd	ShortName	Accper
1	000001	平安银行	2023-12-31
2	000002	万科A	2023-12-31
3	000006	深振业A	2023-12-31
4	000007	*ST 全新	2023-12-31
5	000012	南玻A	2023-12-31
6	000016	深康佳A	2023-12-31
7	000019	深粮控股	2023-12-31
8	000021	深科技	2023-12-31
9	000023	ST 深天	2023-12-31
10	000026	飞亚达	2023-12-31
11	000027	深圳能源	2023-12-31
12	000028	国药一致	2023-12-31
13	000029	深深房A	2023-12-31
14	000031	大悦城	2023-12-31
15	000032	深桑达A	2023-12-31
16	000039	中集集团	2023-12-31
17	000042	中洲控股	2023-12-31
18	000045	深纺织A	2023-12-31
19	000050	深天马A	2023-12-31
20	000055	方大集团	2023-12-31
21	000060	中金岭南	2023-12-31
22	000061	农产品	2023-12-31
23	000062	深圳华强	2023-12-31
24	000063	中兴通讯	2023-12-31
25	000065	北方国际	2023-12-31
26	000066	中国长城	2023-12-31
27	000069	华侨城A	2023-12-31
28	000078	海王生物	2023-12-31
29	000088	盐田港	2023-12-31
30	000089	深圳机场	2023-12-31



	Stkcd	ShortName	Accper
1	000001	平安银行	20231231
2	000002	万科A	20231231
3	000006	深振业A	20231231
4	000007	*ST 全新	20231231
5	000012	南玻A	20231231
6	000016	深康佳A	20231231
7	000019	深粮控股	20231231
8	000021	深科技	20231231
9	000023	ST 深天	20231231
10	000026	飞亚达	20231231
11	000027	深圳能源	20231231
12	000028	国药一致	20231231
13	000029	深深房A	20231231
14	000031	大悦城	20231231
15	000032	深桑达A	20231231
16	000039	中集集团	20231231
17	000042	中洲控股	20231231
18	000045	深纺织A	20231231
19	000050	深天马A	20231231
20	000055	方大集团	20231231
21	000060	中金岭南	20231231
22	000061	农产品	20231231
23	000062	深圳华强	20231231
24	000063	中兴通讯	20231231
25	000065	北方国际	20231231
26	000066	中国长城	20231231
27	000069	华侨城A	20231231
28	000078	海王生物	20231231
29	000088	盐田港	20231231
30	000089	深圳机场	20231231

1.4 substr & substr (第5.1.2题)

```
clear all
input str20 SexAge_Race
"MALE41.2_White"
"FEMALE42.9_White"
"FEMALE38.5_White"
"FEMALE35.6_Hispanic"
"FEMALE48.5_White"
endlist
```

*- Q: 将变量SexAge_Race拆分成三个变量, Sex/Age/Race

1.5 数据合并

```
. merge 1:1 id using filename
```

in memory
master

id	age
1	22
2	56
5	17

in filename.dta
using

id	wgt
1	130
2	180
4	110

= merged result

id	age	wgt
1	22	130
2	56	180
5	17	.
4	.	110

(matched)
(matched)
(master only)
(using only)

Result

Number of obs

Not matched	91
from master	40
from using	51
Matched	40,815

CSDN @Cliche_99

	Stkcd	Accper	Typrep	profit	year	month	asset	_merge
1	000001	2000-12-31	A	5.066e+08	2000	12	.	Master only (1)
2	000001	2001-12-31	A	4.024e+08	2001	12	.	Master only (1)

CSDN @Cliche_99

	Stkcd	Accper	Typrep	profit	year	month	asset	_merge
40902	000007	2015-12-31	A	.	2015	12	4.804e+08	Using only (2)
40903	000007	2016-12-31	A	.	2016	12	5.144e+08	Using only (2)

CSDN @Cliche_99

	Stkcd	Accper	Typrep	profit	year	month	asset	_merge
425	000028	2017-12-31	A	1.157e+09	2017	12	2.234e+10	Matched (3)
426	000028	2018-12-31	A	1.348e+09	2018	12	2.893e+10	Matched (3)

CSDN @Cliche_99

. merge 1:1 pid time using filename

master + *using* = *merged result*

pid	time	x1
14	1	0
14	2	0
14	4	0
16	1	1
16	2	1
17	1	0

pid	time	x2
14	1	7
14	2	9
16	1	2
16	2	3
17	1	5
17	2	2

pid	time	x1	x2	_merge
14	1	0	7	3
14	2	0	9	3
14	4	0	.	1
16	1	1	2	3
16	2	1	3	3
17	1	0	5	3
17	2	.	2	2

. merge 1:m region using filename

master + *using* = *merged result*

region	x
1	15
2	13
3	12
4	11

id	region	a
1	2	26
2	1	29
3	2	22
4	3	21
5	1	24
6	5	20

region	x	id	a	_merge
1	15	2	29	3
1	15	5	24	3
2	13	1	26	3
2	13	3	22	3
3	12	4	21	3
4	11	.	.	1
5	.	6	20	2

2.1 重复值

- duplicates drop ?
- force?
- 练习：
 - (1)去除完全相同的观测值
 - (2)指定的变量有重复，删除这些观测值

2.2 egen

*- (一) 中心化

```
qui sum x  
replace x = x - r(mean)
```

*- (二) 标准化

*- 方法1 :

```
qui sum x  
replace x = (x - r(mean))/r(sd)
```

*- 方法2 :

```
egen z_x = std(x)
```

*- (三) 归一化

```
qui sum x  
replace x = (x - r(min))/(r(max) - r(min))
```

*- 分组中心化、标准化和归一化

```
bysort y: egen max_x = max(x)  
bysort y: egen min_x = min(x)  
bysort y: egen mean_x = mean(x)  
bysort y: egen sd_x = sd(x)
```

```
gen c_x = x - mean_x  
gen z_x = (x - mean_x)/sd_x  
gen n_x = (x - min_x)/(max_x - min_x)
```

*- 太麻烦了，简单方法？

Stata命令运行结果的存储与调用

- r类命令

- 所有其他命令

use grilic.dta,clear

sum s

return list

scalars:

```
r(N) = 758
r(sum_w) = 758
r(mean) = 13.40501319261214
r(Var) = 4.981058057949899
r(sd) = 2.231828411403955
r(min) = 9
r(max) = 18
r(sum) = 10161
```

- e类命令

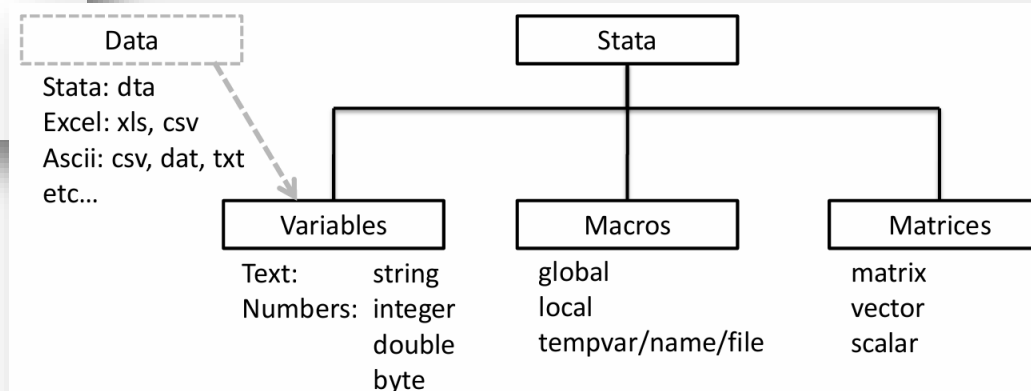
- 估计命令

reg lnw s

ereturn list

四类命令的返回值

三种存储对象



scalars:

```
e(N) = 758
e(df_m) = 1
e(df_r) = 756
e(F) = 255.7039662336329
e(r2) = .2527458374860054
e(rmse) = .3710455612613365
e(mss) = 35.20399459202199
e(rss) = 104.0821552499956
e(r2_a) = .2517574060541087
e(ll) = -323.0498302841153
e(ll_0) = -433.4714451849197
e(rank) = 2
```

macros:

```
e(cmdline) : "regress lnw s"
e(title) : "Linear regression"
e(marginsok) : "XB default"
e(vce) : "ols"
e(depvar) : "lnw"
e(cmd) : "regress"
e(properties) : "b V"
e(predict) : "regres_p"
e(model) : "ols"
e(estat_cmd) : "regress_estat"
```

matrices:

```
e(b) : 1 x 2
e(V) : 2 x 2
```

functions:

```
e(sample)
```

2.5 描述性统计练习

Table 11: Summary statistics for the slave trade and trust variables.

	Summary Statistics				
	Obs.	Mean	S.d.	Min.	Max.
<u>Trust variables:</u>					
Trust relatives	20,618	2.19	0.96	0	3
Trust neighbors	20,580	1.74	1.01	0	3
Trust local council	20,210	1.66	1.10	0	3
Intra-group trust	20,502	1.67	1.00	0	3
Inter-group trust	20,301	1.36	1.00	0	3
<u>Slave trade variables:</u>					
slave exports (thousands)	21,702	93.17	205.28	0	854.96
ln (1 + slave exports)	21,702	1.95	2.31	0	6.75
exports/area	21,702	2.66	7.68	0	37.70
exports/historic pop	18,366	0.11	0.22	0	4.46
ln(1+exports/area)	21,702	0.53	0.95	0	3.66
ln(1+exports/historic pop)	18,366	0.09	0.17	0	1.70

导出Table 11的Stata命令为?

作答

练习

1. 数据准备和基本操作

1.1 加载数据和初步探索

1.2 数据管理和转换

2. 描述性统计和数据可视化

2.1 基本描述性统计

2.2 数据可视化

3. 假设检验

4. 回归分析

4.1 基本回归分析

4.2 回归诊断

5. 高级分析技术

5.1 非线性关系分析

5.2 模型选择和诊断

5.3 分位数回归

5.4 内生性和工具变量

5.5 中介分析

5.6 交互效应分析

5.7 聚类分析

6. 结果导出

练习

- 加载auto数据集
- 查看数据结构
- 查看前5行数据
- 查看变量标签
- 查看值标签
- 浏览数据

练习

- 生成对数价格变量
- 生成哑变量：是否是美国车
- 将mpg转换为升/100公里
- 对连续变量weight进行中心化
- 对price按照分位数进行分类
- 标记price的异常值（比如，超过3个标准差的观察值）
- 对price进行排序
- 保留需要的变量

练习

- 基本描述性统计
- 按照原产地分组的描述性统计
- 相关性分析

练习

- 绘制散点图矩阵
- 绘制箱线图
- 绘制核密度图
- 绘制直方图
- 绘制散点图并添加拟合线
- 绘制条形图

练习

- 进行t检验，比较国产车和进口车的价格差异
- 进行方差分析 (ANOVA)
- 进行卡方检验，检验车型和原产地是否独立

练习

- 简单线性回归
- 多元线性回归
- 加入交互项的回归
- 加入分类变量的回归
- 稳健标准误

练习

- 检查多重共线性
- 异方差性检验
- 模型诊断图

练习

- 探索价格和重量之间的非线性关系
- 多项式回归
- 对数-对数模型
- 半对数模型
- 反向回归

练习

• 逐步回归

Stata实现逐步回归法

向前逐步回归Forward selection:

```
stepwise regress y x1 x2 ... xk, pe(#1)
```

pe(#1) specifies the significance level for addition to the model; terms with $p < \#1$ are eligible for addition (显著才加入模型中)。

向后逐步回归Backward elimination:

```
stepwise regress y x1 x2 ... xk, pr(#2)
```

pr(#2) specifies the significance level for removal from the model; terms with $p \geq \#2$ are eligible for removal (不显著就剔除出模型)。

如果你觉得筛选后的变量仍很多，你可以减小#1或者#2

如果你觉得筛选后的变量太少了，你可以增加#1或者#2

注:

- (1) $x_1 x_2 \dots x_k$ 之间不能有完全多重共线性(和regress不同哦)
- (2) 可以在后面再加参数b和r，即标准化回归系数或稳健标准误

CSDN @疯狂成露者

逐步回归分析

向前逐步回归Forward selection: 将自变量逐个引入模型，每引入一个自变量后都要进行检验，显著时才加入回归模型。

(缺点：随着以后其他自变量的引入，原来显著的自变量也可能又变为不显著了，但是，并没有将其及时从回归方程中剔除掉。)

向后逐步回归Backward elimination: 与向前逐步回归相反，先将所有变量均放入模型，之后尝试将其中一个自变量从模型中剔除，看整个模型解释因变量的变异是否有显著变化，之后将最没有解释力的那个自变量剔除；此过程不断迭代，直到没有自变量符合剔除的条件。(缺点：一开始把全部变量都引入回归方程，这样计算量比较大。若对一些不重要的变量，一开始就不引入，这样就可以减少一些计算。当然这个缺点随着现在计算机的能力的提升，已经变得不算问题了)

CSDN @疯狂成露者

逐步回归的说明

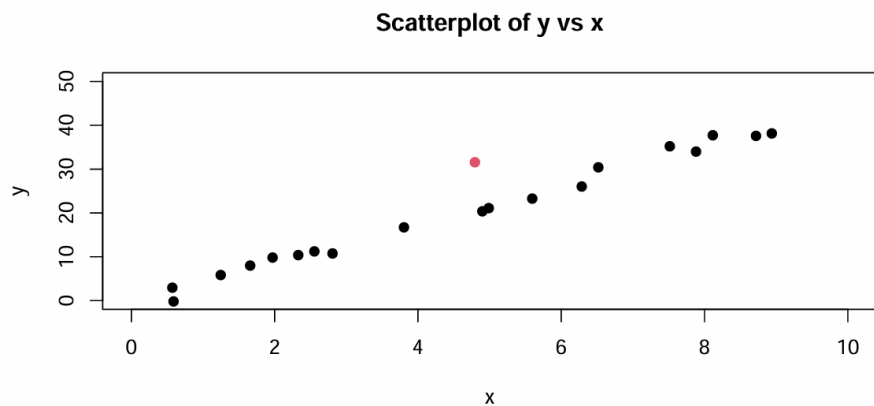
- (1) 向前逐步回归和向后逐步回归的结果可能不同。
- (2) 不要轻易使用逐步回归分析，因为剔除了自变量后很有可能会产生新的问题，例如内生性问题。
- (3) 有没有更加优秀的筛选方法？有的，那就是每种情况都尝试一次，最终一共有 $C_k^1 + C_k^2 + \dots + C_k^k = 2^k - 1$ 种可能。如果自变量很多，那么计算相当费时。

CSDN @疯狂成露者

练习

- Ramsey RESET检验
- 检查异常值和高杠杆值点

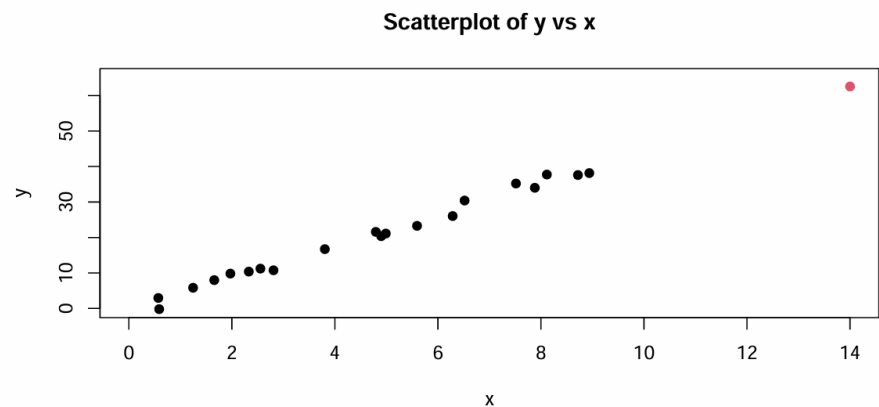
Leverage vs. Outliers example



The red data point is outlying, but not influential. How can you check this?

CSDN @冰霜青羽

Leverage vs. Outliers example



The red data point is not an outlier, it has leverage, but it is not influential.

CSDN @冰霜青羽

练习

- 探索重量和原产地对价格的交互作用
- 可视化交互效应

练习

- 运行几个回归并将结果导出到Word

3.2 取整和四舍五入

取整

```
. di int(3.49)           //int()取整, 不论后面的小数是什么, 只取小数点前的数值
. di int(3.51)           //输出 3
. di int(-3.49)          //输出 -3
. di int(-3.51)          //输出 -3
```

四舍五入

```
. di round(3.49)         //round()取整, 四舍五入, 结果为 3
. di round(3.51)         //四舍五入, 结果为 4
. di round(-3.49)        //四舍五入, 结果为 -3
. di round(-3.51)        //四舍五入到个位数, 结果为 -4

. di round(3.345,.1)     //四舍五入到十分位, 结果为 3.3
. di round(3.351,.1)     //四舍五入到十分位, 结果为 3.4
. di round(3.345,.01)    //四舍五入到百分位, 结果为 3.35
. di round(3.351,.01)    //四舍五入到百分位, 结果为 3.35
. di round(335.1,10)     //四舍五入到十位, 结果为 340
```

3.6 回归

表 6.1 回归中 options 的内容表

<code>noconstant</code>	模型不包含常数项
<code>hascons</code>	用户自定义常数项
<code>level(#)</code>	设置置信区间，默认值 95%
<code>beta</code>	标准化系数
<code>vce(type)</code>	设置估计量的标准差，常用的主要有： <code>ols</code> ， <code>robust</code> ， <code>bootstrap</code> ， <code>hc2</code> ，等

`vce(hc2)` and specify an alternative bias correction for the robust variance calculation. and may not be specified with the prefix. In the unclustered case, uses $(\hat{\sigma}_j)^2 = \{n/(n-k)\}(u_j)^2$ as an estimate of the variance of the j th observation, where u_j is the calculated residual and $n/(n-k)$ is included to improve the overall estimate's small-sample properties.

`vce(hc3)`
`vce(hc2)`
`vce(hc3)`
`svy`
`vce(robust)`

`vce(hc2)` instead uses $u_j^2/(1-h_{jj})$ as the observation's variance estimate, where h_{jj} is the diagonal element of the hat (projection) matrix. This estimate is unbiased if the model really is homoskedastic. tends to produce slightly more conservative confidence intervals. `vce(hc2)`

`vce(hc3)` uses $u_j^2/(1-h_{jj})^2$ as suggested by [Davidson and MacKinnon \(1993\)](#), who report that this method tends to produce better results when the model really is heteroskedastic. produces confidence intervals that tend to be even more conservative.

`vce(hc3)`

See [Davidson and MacKinnon \(1993, 554-556\)](#) and [Angrist and Pischke \(2009, 294-308\)](#) for more discussion on these two bias corrections.

3.7 偏差方向

- 是看系数的正负号吗？

存在严重的多重共线性时，参数估计值的标准差（）

- ☒ A 变大
- ☐ B 变小
- ☐ C 不变
- ☐ D 无法估计

提交

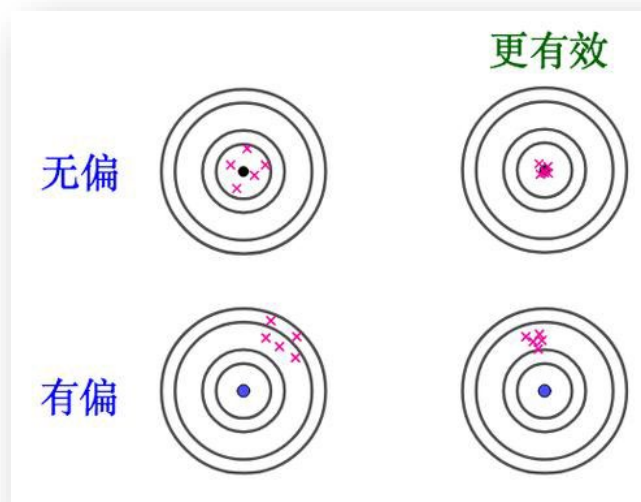
3.8 正态性检验

01	图形法	histogram var1, normal
02	正态概率图 (Q-Q 图)	qnorm var1
03	夏皮罗 - 威尔克 (Shapiro - Wilk) 检验	swilk var1
04	柯尔莫哥洛夫 - 斯米尔诺夫 (Kolmogorov - Smirnov) 检验	sktest var1
05	雅克 - 贝拉 (Jarque - Bera) 检验	ssc install jb jb var1

3.9 内生性

- 内生性的来源

- 遗漏变量
- 选择偏差
 - 样本选择偏差
 - 自选择偏差
- 双向因果
- 测量误差



- 内生性的解决

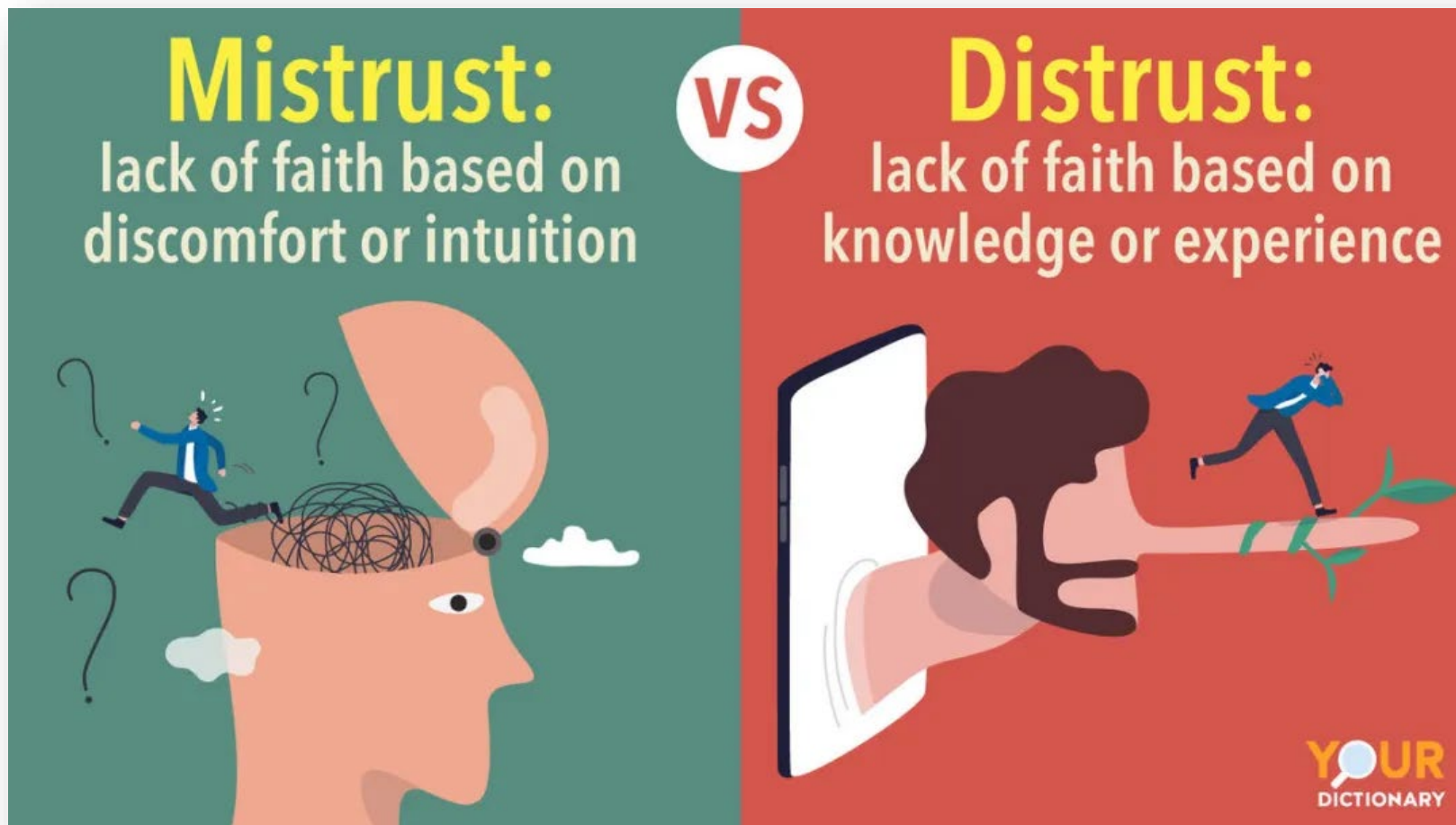
- 工具变量法
- 自然实验法
- 处理效应模型
- Heckman模型
- 引入固定效应
- 广义矩估计
- 断点回归
- 倾向得分匹配模型

3.10 预测

表 6.2 预测中 options 的内容表

<u>xb</u>	线性预测拟合值
residuals 或者 score	残差
<u>rstandard</u>	标准化的残差
<u>rstudent</u>	学生化的残差
<u>stdp</u>	样本内预测标准差
<u>stdf</u>	样本外预测标准差
<u>stdr</u>	残差的标准差
<u>cooks</u>	Cook 的 D 影响统计量
<u>covratio</u>	COVATIO 影响统计量
<u>dfits</u>	DFITS 影响统计量
<u>welsch</u>	Welsh 距离
<u>dfbeta(varname)</u>	变量 <u>varname</u> 的 DFBETA

4 经济史中的数据



回归结果

TABLE 1—OLS ESTIMATES OF THE DETERMINANTS OF TRUST IN NEIGHBORS

Dependent variable: Trust of neighbors	Slave exports (thousands) (1)	Exports/ area (2)	Exports/ historical pop (3)	ln (1 + exports) (4)	ln (1 + exports/ area) (5)	ln (1 + exports/ historical pop) (6)
Estimated coefficient	−0.00068 [0.00014] (0.00015) {0.00013}	−0.019 [0.005] (0.005) {0.005}	−0.531 [0.147] (0.147) {0.165}	−0.037 [0.014] (0.014) {0.015}	−0.159 [0.034] (0.034) {0.034}	−0.743 [0.187] (0.187) {0.212}
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	20,027	20,027	17,644	20,027	20,027	17,644
Number of ethnicities	185	185	157	185	185	157
Number of districts	1,257	1,257	1,214	1,257	1,257	1,214
R^2	0.16	0.16	0.15	0.15	0.16	0.15

Notes: The table reports OLS estimates. The unit of observation is an individual. Below each coefficient three standard errors are reported. The first, reported in square brackets, is standard errors adjusted for clustering within ethnic groups. The second, reported in parentheses, is standard errors adjusted for two-way clustering within ethnic groups and within districts. The third, reported in curly brackets, is T. G. Conley (1999) standard errors adjusted for two-dimensional spatial autocorrelation. The standard errors are constructed assuming a window with weights equal to one for observations less than five degrees apart and zero for observations further apart. The individual controls are for age, age squared, a gender indicator variable, five living conditions fixed effects, ten education fixed effects, 18 religion fixed effects, 25 occupation fixed effects, and an indicator for whether the respondent lives in an urban location. The district controls include ethnic fractionalization of each district and the share of the district's population that is the same ethnicity as the respondent.

```

16 *****
17 *** Table 1: Trust Neighbors ***
18 *****
19
20 *** Slave exports only ***
21 xi: reg trust_neighbors exports `baseline_controls', cluster(murdock_name)
22 xi: cgmreg trust_neighbors exports `baseline_controls', cluster(murdock_name district)
23 *** Slave exports/land area ***
24 xi: reg trust_neighbors export_area `baseline_controls', cluster(murdock_name)
25 xi: cgmreg trust_neighbors export_area `baseline_controls', cluster(murdock_name district)
26 *** slave exports/historic pop ***
27 xi: reg trust_neighbors export_pop `baseline_controls', cluster(murdock_name)
28 xi: cgmreg trust_neighbors export_pop `baseline_controls', cluster(murdock_name district)
29 *** ln slave exports ***
30 xi: reg trust_neighbors ln_exports `baseline_controls', cluster(murdock_name)
31 xi: cgmreg trust_neighbors ln_exports `baseline_controls', cluster(murdock_name district)
32 *** ln slave exports/land area ***
33 xi: reg trust_neighbors ln_export_area `baseline_controls', cluster(murdock_name)
34 xi: cgmreg trust_neighbors ln_export_area `baseline_controls', cluster(murdock_name district)
35 *** ln slave exports/historic pop ***
36 xi: reg trust_neighbors ln_export_pop `baseline_controls', cluster(murdock_name)
37 xi: cgmreg trust_neighbors ln_export_pop `baseline_controls', cluster(murdock_name district)
38
39 /* Regressions to get number of clusters info (since CGM ado file does not handle this properly) */
40 *** Slave exports only ***
41 xi: reg trust_neighbors exports `baseline_controls', cluster(murdock_name)
42 xi: reg trust_neighbors exports `baseline_controls', cluster(district)
43 *** Slave exports/land area ***
44 xi: reg trust_neighbors export_area `baseline_controls', cluster(murdock_name)
45 xi: reg trust_neighbors export_area `baseline_controls', cluster(district)
46 *** slave exports/historic pop ***
47 xi: reg trust_neighbors export_pop `baseline_controls', cluster(murdock_name)
48 xi: reg trust_neighbors export_pop `baseline_controls', cluster(district)
49 *** ln slave exports ***
50 xi: reg trust_neighbors ln_exports `baseline_controls', cluster(murdock_name)
51 xi: reg trust_neighbors ln_exports `baseline_controls', cluster(district)
52 *** ln slave exports/land area ***
53 xi: reg trust_neighbors ln_export_area `baseline_controls', cluster(murdock_name)
54 xi: reg trust_neighbors ln_export_area `baseline_controls', cluster(district)
55 *** ln slave exports/historic pop ***
56 xi: reg trust_neighbors ln_export_pop `baseline_controls', cluster(murdock_name)
57 xi: reg trust_neighbors ln_export_pop `baseline_controls', cluster(district)

```

修改：使用循环语句简化

```
1  local i = 1
2  foreach v1 in trust_neighbors{
3      local j = 1
4      foreach v2 in $depend_var{
5          qui xi: reg `v1' `v2' $baseline_controls, cluster(murdock_name)
6          eststo m1_`i'_`j'
7
8          qui xi: cgmreg `v1' `v2' $baseline_controls, cluster(murdock_name district)
9          eststo m2_`i'_`j'
10
11         qui xi: reg `v1' `v2' $baseline_controls, cluster(district)
12         eststo m4_`i'_`j'
13
14         local j = `j' + 1
15     }
16     local i = `i' + 1
17 }
18 esttab m*
```

```

60 *****
61 *** Table 2: All Trust Measures ***
62 *****
63
64 *** Relatives Trust ***
65 xi: cgmreg trust_relatives ln_export_area `baseline_controls', cluster(murdock_name district)
66 *** Neighbors Trust ***
67 xi: cgmreg trust_neighbors ln_export_area `baseline_controls', cluster(murdock_name district)
68 *** Trust Local Council ***
69 xi: cgmreg trust_local_council ln_export_area `baseline_controls', cluster(murdock_name district)
70 *** Intra Group Trust ***
71 xi: cgmreg intra_group_trust ln_export_area `baseline_controls', cluster(murdock_name district)
72 *** Inter Group Trust ***
73 xi: cgmreg inter_group_trust ln_export_area `baseline_controls', cluster(murdock_name district)
74
75 /* Regressions to get number of clusters info (since CGM ado file does not handle this properly) */
76 *** Relatives Trust ***
77 xi: reg trust_relatives ln_export_area `baseline_controls', cluster(murdock_name)
78 xi: reg trust_relatives ln_export_area `baseline_controls', cluster(district)
79 *** Neighbors Trust ***
80 xi: reg trust_neighbors ln_export_area `baseline_controls', cluster(murdock_name)
81 xi: reg trust_neighbors ln_export_area `baseline_controls', cluster(district)
82 *** Trust Local Council ***
83 xi: reg trust_local_council ln_export_area `baseline_controls', cluster(murdock_name)
84 xi: reg trust_local_council ln_export_area `baseline_controls', cluster(district)
85 *** Intra Group Trust ***
86 xi: reg intra_group_trust ln_export_area `baseline_controls', cluster(murdock_name)
87 xi: reg intra_group_trust ln_export_area `baseline_controls', cluster(district)
88 *** Inter Group Trust ***
89 xi: reg inter_group_trust ln_export_area `baseline_controls', cluster(murdock_name)
90 xi: reg inter_group_trust ln_export_area `baseline_controls', cluster(district)

```

```

local i = 1
local j = 1
foreach v1 in $depend_var{
    local j = 1
    foreach v2 of varlist ln_export_area{
        qui xi: reg `v1' `v2' $baseline_controls, cluster(murdock)
        estadd local Individual_controls "Yes", replace
        estadd local District_controls "Yes", replace
        estadd local Country_fixed_effects "Yes", replace
        eststo _m1_`i'`j'
        local j = `j' + 1
    }
    local i = `i' + 1
}
esttab _m*, b(%8.3f) se(%8.3f) keep(ln_export_area) nogap stat(Individual_controls District_controls Country_fixed_effects N r2, fmt(%3s %3s %3s %12.0f %9.2f))

```

TABLE 2—OLS ESTIMATES OF THE DETERMINANTS OF THE TRUST OF OTHERS

	Trust of relatives (1)	Trust of neighbors (2)	Trust of local council (3)	Intra- group trust (4)	Inter- group trust (5)
ln (1 + exports/area)	-0.133*** (0.037)	-0.159*** (0.034)	-0.111*** (0.021)	-0.144*** (0.032)	-0.097*** (0.028)
Individual controls	Yes	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Number of observations	20,062	20,027	19,733	19,952	19,765
Number of ethnicity clusters	185	185	185	185	185
Number of district clusters	1,257	1,257	1,283	1,257	1,255
R ²	0.13	0.16	0.20	0.14	0.11

	(1) trust_rela~s	(2) trust_neig~s	(3) trust_loca~l	(4) trust_intr~p	(5) trust_inte~p
ln_export~a	-0.133*** (0.036)	-0.159*** (0.034)	-0.111*** (0.022)	-0.144*** (0.031)	-0.097*** (0.028)
Individual~s	Yes	Yes	Yes	Yes	Yes
District_c~s	Yes	Yes	Yes	Yes	Yes
Country_fi~s	Yes	Yes	Yes	Yes	Yes
N	20062	20027	19733	19952	19765
r2	0.13	0.16	0.20	0.14	0.11
Standard errors in parentheses * p<0.05, ** p<0.01, *** p<0.001					*/

因果推断经验研究中的中介效应与调节效应

江 艇

摘要 中介效应分析与调节效应分析是国内经济学因果推断经验研究中被广泛采用的两种研究手段,但是存在不同程度的误用。前者的主要问题在于过度使用中介效应逐步法检验,后者主要问题是对其在因果识别中的作用阐发不足。为此,本文深入讨论了中介效应检验的偏误、如何正确开展中介效应分析,以及如何使用调节效应分析来强化因果关系论证,并针对当前的使用现状提出了相应的操作建议。

关键词 因果推断; 中介效应; 调节效应

[中图分类号] F064 **[文献标识码]** A **[文章编号]** 1006-480X(2022)05-0100-21

DOI:10.19581/j.cnki.ciejournal.2022.05.005

Nunn and Wantchekon(2011) 使用 2005 年非洲民意调查数据,检验非洲历史上的奴隶贸易造成的人际不信任是否延续至今。被解释变量是受访者报告的各种信任指标,反映受访者是否愿意信任他人,核心解释变量是受访者所在种族历史上的奴隶出口数量。结果显示,历史上奴隶出口越猖獗,如今人们就越不愿意信任他人。在“因果渠道检验”(Testing for Channels of Causality)中指出,奴隶贸易影响的可能是人们的内心行为规范(不愿意信任他人),也可能是外部环境(由于国家、制度和法律的衰败,他人不再值得信任)。他们想要强调的是前一种因果渠道,于是构造了能够反映后一种竞争性解释的指标并在回归中予以控制,比较控制前后核心解释变量的系数变化,来看奴隶贸易的效应是否被竞争性解释所吸收。例如,控制反映政府质量的变量(此时以对当地政府的信任作为被解释变量),奴隶贸易的系数变小了 50%,这说明至少有一半的效应无法被外部环境的恶化所解释。又如,控制受访者所在地其他种族历史上受奴隶贸易的影响,从而间接控制周围他人的可信任程度(此时用族群间信任作为被解释变量),奴隶贸易的系数基本没有变化,说明奴隶贸易对今天族群间信任的效应无法被解释成奴隶贸易降低了他人的可信任程度,而只能解释为奴隶贸易使人更不愿意信任他人。

What is a mechanism?

The purpose of this book is to help you learn how to draw reasonable mechanisms for organic reactions. A *mechanism* is a story that we tell to explain how compound **A** is transformed into compound **B** under given reaction conditions. Imagine being asked to describe how you traveled from New York to Los Angeles (an overall reaction). You might tell how you traveled through New Jersey to Pennsylvania, across to St. Louis, over to Denver, then through the Southwest to the West Coast (the mechanism). You might include details about the mode of transportation you used (reaction conditions), cities where you stopped for a few days (intermediates), detours you took (side reactions), and your speed at various points along the route (rates). To carry the analogy further, there is more than one way to get from New York to Los Angeles; at the same time, not every story about how you traveled from New York to Los Angeles is believable. Likewise, more than one reasonable mechanism can often be drawn for a reaction, and one of the purposes of this book is to teach you how to distinguish a reasonable mechanism from a whopper.

期末作业：论文复现要求

- 自选一篇文章进行阅读和复现
- 文献选取要求
 - 英文期刊在ABS 3星及以上
 - 中文期刊至少是南大核心
- 需要讲解各部分Stata代码的含义，建议时长8-10分钟

注意事项

- 在选择论文时，请尤其注意数据可获得性，请勿选择使用非公开数据的论文。
- 建议使用Stata进行实证分析。
- 复现内容需包括：
 - 描述性统计
 - 基准回归结果
 - 主要稳健性结果
 -
- 并需与原文结果进行对比，若无法复现原文结果，需标记说明并尝试解释。