Universidade de São Paulo / Faculdade de Filosofia, Letras e Ciências Humanas
Departamento de Ciência Política
FLP-406 & FLS-6183
2º semestre / 2017

Lista 9

Neste laboratório iremos nos concentrar em aprofundar as noções básicas de interação na regressão multivariada, em especial, utilizando o pacote ℭlarify de Tomz, Wittenberg & King (2003) para *Stata*. Vamos comparar os resultados de Clarify com os resultados obtidos no laboratório 3 realizado na semana passada.

Parte I – Sintaxe Básica do ℭlarify: Simulações, Níveis e Quantidades de Interesse

O ℭlarify é um pacote de ferramentas estatísticas desenvolvido com o fim de produzir resultados estatísticos que sejam do interesse do pesquisador e, ao mesmo tempo, de fácil consumo pelo público não especializado. Uma característica importante do pacote é que as estimativas são geradas por meio de simulações aleatórias, com a consequência de que nem sempre um mesmo *script* produzirá os mesmos resultados.

Por ser um pacote desenvolvido por terceiros, o ℭlarify deve ser instalado separadamente. Primeiramente, digite na linha de comando no Stata:

```
net from http://gking.harvard.edu/clarify
```

Em seguida, digite o seguinte comando:

```
net install clarify
```

Pronto! O pacote está instalado e já pode ser utilizado.

Para entender o funcionamento do pacote, continuaremos utilizando o banco de dados `Coxappend.dta` para avaliar a teoria de Duverger (1954).

As rotinas em ℭlarify consistem em basicamente três comandos (e suas diversas opções):

- `-estsimp-` : gera valores aleatórios a partir de um modelo estimado (ex: `estsimp regress y x`)
- `-setx-` : nível de *X* a ser fixado para simular valores de *y* (ex: `setx mean`)
- `-simqi-` : simulação de quantidades de *y* com base nos níveis fixados.

Para este exemplo, continuamos utilizando o modelo:

$$\begin{aligned} &LegislativeParties = \\ &= \beta_0 + \beta_1 MultimemberDistrict + \beta_2 SocialHeterogeneity \\ &+ \beta_3 (MultimemberDistrict \times SocialHeterogeneity) + \varepsilon \end{aligned} \qquad \text{(I)}$$

Parte II: Variáveis Dicotômicas

Antes de estimar o modelo, você deve gerar a variável interativa. O ℭlarify não entende a sintaxe de interação do Stata (# ou ##). Mas, no banco de dados `Coxappend.dta`, o termo interativo já foi fornecido. Caso não existisse, uma nova variável deveria ser criada:

```
gen multi_heterogeneity = multimember*heterogeneity
```

Em seguida, estime o modelo com o primeiro comando do ℭlarify:

```
estsimp regress enps heterogeneity multimember multi_heterogeneity
```

Com este primeiro comando, o ℭlarify gera uma série de valores simulados para os betas (`b1`, `b2`, `b3`, `b4`) e para a variância de $\hat{y}$ (`b5`).

```
setx heterogeneity 0 multimember 0 multi_heterogeneity 0
```

O segundo comando fixo os valores de todas as variáveis independentes.

```
simqi, pv
```

Este comando apresentará o valor predito de *y* com *x* tal como especificado no comando anterior de setx.

Com esse primeiro resultado, já temos uma das células da Tabela 1.

Table 1. The Predicted Number of Legislative Parties with Clarify with Dichotomous Variables

Social Heterogeneity	Single-Member Districts	Multi-Member Districts
Heterogeneous		
Homogenous	2.478 [-0.008, 5.065]	

Exercício 1. Agora, você deve criar o código para preencher os valores preditos em os outros três casos.

Exercício 2. Como se comparam estes resultados obtidos com os resultados que foram calculados no laboratório anterior e reportados na Tabela 3 (veja abaixo)?

Table 3. The Predicted Number of Legislative Parties using Lincom Command

Social Heterogeneity	Single-Member Districts	Multi-Member Districts
Heterogeneous	1.68 [0.94 – 2.43]	3.88 [3.26 – 4.50]
Homogenous	2.52 [1.69 – 3.34]	3.06 [2.48 – 3.65]

Notes: 95% confidence intervals in parentheses

Part III. X e Z são variáveis contínuas.

Agora vamos estimar o mesmo modelo com as variáveis dependentes e explicativas em logs. Primeiro, vamos dar um `clear` e voltar a carregar o banco.

Diferente do laboratório da semana passada, vamos agora estimar o modelo com todas as variáveis em logs.

gen lnenps=ln(enps)

gen lneneth=ln(eneth)

```
gen lmleneth2 = lnml * lneneth
```

Exercício 1. O que é revelado sobre as hipóteses pela regressão?

```
. estsimp regress lnenps lneneth lnml lmleneth2
```

Source	SS	df	MS	Number of obs	=	54
				F(3, 50)	=	14.09
Model	4.31454195	3	1.43818065	Prob > F	=	0.0000
Residual	5.10222521	50	.102044504	R-squared	=	0.4582
				Adj R-squared	=	0.4257
Total	9.41676716	53	.177674852	Root MSE	=	.31944

lnenps	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lneneth	-.4179624	.1868026	-2.24	0.030	-.7931665	-.0427583
lnml	.0694294	.0457695	1.52	0.136	-.0225013	.1613601
lmleneth2	.3181527	.0945409	3.37	0.001	.1282617	.5080437
_cons	.865806	.098982	8.75	0.000	.6669947	1.064617

Novamente, agora precisamos pensar em como estes resultados podem ser entendidos.

Vamos pensar nos casos acima especificados continuando com o intento de testar as hipóteses de Duverger. Precisamos de pensar em valores para cada variável explicativa tais que podemos novamente pensar em entender a diferença entre a) single member districts e sociedades homogêneas b) multi-member districts e sociedades homogêneas, c) single member districts e sociedades heterogêneas; e, d) multi-member districts e sociedades heterogêneas,

Nota: Como explicam King, Tomz e Wittenberg (2000), "If some function of Y, such as ln (Y) is used as the dependent variable in the estimation stage, the researcher can simulate ln (Y) and then apply the inverse function exp (ln(y)) to reveal Y.

No caso de single member districts em sociedades homogêneas,

P_{ml}=ln(median district magnitude=1)=0.

P_{hom}=ln (effective number of ethnic groups=25th percentile=1.105)= 0.09984533

P_{ml*hom}=ln ((median dm =1)* (effective number of ethnic groups=25th percentile=1.105))= ln (1.105)= 0.09984533

Desejamos analisar a distribuição da quantidade predita de partidos políticos. Para isso, você necessitará ajustar as opções dos comandos –setx– e –simqi–.

Para avaliar a distribuição de $\hat{y}$ quando há a mediana da magnitude do distrito é 1 e a sociedade é homogênea (25 percentile de ethnic groups):

setx lnml 0 lneneth 0.09984533 lmleneth2 0.09984533

simqi, pv genpv(prevalsmhomog)

(além de gerar os valores criamos uma nova variável de nome "preval_sm_homog" com valor predito de y neste caso.

Os resultados são:

. simqi, pv genpv(preval_sm_homog)

Quantity of Interest	Mean	Std. Err.	[95% Conf. Interval]
Pred(lnenps)	.8549754	.3371482	.1772649 1.528141

Para converter este valor em valores que podemos comparar com nossas variáveis na métrica original, precisamos agora obter:

di exp(0.8549754)

di exp(0.1772649)

di exp (1.528141)

Agora, temos os primeiros resultados para a Tabela 2.

Table 2. The Predicted Number of Legislative Parties with Clarify with Continuous Variables

Social Heterogeneity	Single-Member Districts	Multi-Member Districts
Heterogeneous		
Homogenous	2.3513165 [1.1939473, 4.6095996]	

Exercício 2. Agora, você deve criar o código para preencher os valores preditos nos outros três casos.

Exercício 3. Como se comparam estes resultados obtidos com os resultados que foram calculados no laboratório anterior e reportados na Tabela 3 ?