```
                                                                  ___  ____  ____  ____  ____(R)
                                                                 /__    /   ____/   /   ____/
                                                                ___/   /   /___/   /   /___/
                                                                  Statistics/Data analysis

                                                                  User: REBECA SILVA BARROSO
                                Project: ANALISE DA TAXA DE ANALFABETISMO EM ALAGOAS NO PERÍODO DE 2000 E 2010

  ___  ____  ____  ____  ____ ©
 /__    /   ____/   /   ____/      17.0
___/   /   /___/   /   /___/       MP—Parallel Edition

 Statistics and Data Science       Copyright 1985-2021 StataCorp LLC
                                   StataCorp
                                   4905 Lakeway Drive
                                   College Station, Texas 77845 USA
                                   800-STATA-PC        https://www.stata.com
                                   979-696-4600        stata@stata.com

Stata license: Unlimited-user 64-core network perpetual
Serial number: 18461036
  Licensed to: TEAM BTCR
               TEAM BTCR

Notes:
      1. Unicode is supported; see help unicode_advice.
      2. More than 2 billion observations are allowed; see help obs_advice.
      3. Maximum number of variables is set to 5,000; see help set_maxvar.

 1 . do "C:\Users\rebec\OneDrive\Mestrado\Disciplinas\2023.2\Econometria\Trabalho final\script.do"

 2 . ***** trabalho econometria *****
 3 .
 4 . ***** importando a base de dados *****
 5 .
 6 . import excel "C:\Users\rebec\OneDrive\Mestrado\Disciplinas\2023.2\Econometria\Trabalho final\base de dados.xlsx",
   > irstrow
   (11 vars, 204 obs)

 7 .
 8 . * mostrando ao stata os anos da base *
 9 .
10 . xtset cod ano, yearly

   Panel variable: cod (strongly balanced)
    Time variable: ano, 2000 to 2010, but with gaps
            Delta: 1 year

11 .
12 . *** Examinando o banco de dados *****
13 . describe

   Contains data
    Observations:           204
       Variables:            11
   ------------------------------------------------------------------------------------------------------------------------
   Variable      Storage   Display    Value
       name         type    format    label      Variable label
   ------------------------------------------------------------------------------------------------------------------------
   cod             int     %10.0g                cod
   mun             str29   %29s                  mun
   ano             int     %ty                   ano
   analf           double  %10.0g                analf
   theil           double  %10.0g                theil
   gini            double  %10.0g                gini
   dens            double  %10.0g                dens
   idhm            double  %10.0g                idhm
   idhm_educ       double  %10.0g                idhm_educ
   ensino_medio    double  %10.0g                ensino_medio
   renda           double  %10.0g                renda
   ------------------------------------------------------------------------------------------------------------------------
   Sorted by: cod  ano
        Note: Dataset has changed since last saved.
```

```
14 . list in 1/5
```

	cod	mun	ano	analf	theil	gini	dens	idhm	idhm_e~c	ensino~o	renda
1.	1	Água Branca (AL)	2000	43.7	.72	.66	59.07	.392	.195	7.85	138.86
2.	1	Água Branca (AL)	2010	33.78	.63	.56	40.81	.549	.432	15.77	213
3.	2	Anadia (AL)	2000	45.52	.55	.58	52.81	.428	.242	8.16	170.3
4.	2	Anadia (AL)	2010	39.1	.52	.53	33.99	.568	.444	17.66	239.53
5.	3	Arapiraca (AL)	2000	32.45	.57	.57	43.96	.476	.297	12.44	258.45

```
15 .
16 . **** Análise descritiva dos dados *************
17 . summ analf gini dens idhm ensino_medio renda
```

Variable	Obs	Mean	Std. dev.	Min	Max
analf	204	40.71294	8.835575	12.51	57.79
gini	204	.5584804	.0661625	.41	.82
dens	204	47.21309	10.44164	25.12	73.15
idhm	204	.4774265	.0975803	.281	.721
ensino_medio	204	11.76961	6.625598	1.29	43.16
renda	204	202.6036	86.02629	69.2	792.54

```
18 . summ analf gini dens idhm ensino_medio renda, detail
```

analf

Percentiles		Smallest		
1%	19.57	12.51		
5%	25.59	17.68		
10%	30.26	19.57	Obs	204
25%	34.305	19.64	Sum of wgt.	204
50%	41.255		Mean	40.71294
		Largest	Std. dev.	8.835575
75%	47.01	56.48		
90%	52.85	56.94	Variance	78.06739
95%	54.04	57.34	Skewness	-.2898274
99%	56.94	57.79	Kurtosis	2.765881

gini

Percentiles		Smallest		
1%	.43	.41		
5%	.47	.42		
10%	.48	.43	Obs	204
25%	.51	.43	Sum of wgt.	204
50%	.55		Mean	.5584804
		Largest	Std. dev.	.0661625
75%	.6	.71		
90%	.65	.72	Variance	.0043775
95%	.67	.75	Skewness	.6334702
99%	.72	.82	Kurtosis	3.604016

dens

Percentiles		Smallest		
1%	26.45	25.12		
5%	30.83	25.86		
10%	33.62	26.45	Obs	204
25%	39.27	26.96	Sum of wgt.	204
50%	47.045		Mean	47.21309
		Largest	Std. dev.	10.44164
75%	55.64	65.51		
90%	60.85	65.78	Variance	109.0279
95%	63.43	66.36	Skewness	-.0410731
99%	65.78	73.15	Kurtosis	2.078697

idhm

Percentiles		Smallest		
1%	.301	.281		
5%	.335	.299		
10%	.354	.301	Obs	204
25%	.3845	.306	Sum of wgt.	204
50%	.5045		Mean	.4774265
		Largest	Std. dev.	.0975803
75%	.5635	.643		
90%	.591	.649	Variance	.0095219
95%	.612	.66	Skewness	-.059301
99%	.649	.721	Kurtosis	1.766908

ensino_medio

Percentiles		Smallest		
1%	2.06	1.29		
5%	3.27	1.73		
10%	4.28	2.06	Obs	204
25%	6.635	2.26	Sum of wgt.	204
50%	10.375		Mean	11.76961
		Largest	Std. dev.	6.625598
75%	15.825	27.98		
90%	20.4	29.64	Variance	43.89854
95%	23.02	33.33	Skewness	1.007633
99%	29.64	43.16	Kurtosis	4.829557

renda

Percentiles		Smallest		
1%	82.3	69.2		
5%	99.88	79.05		
10%	118.77	82.3	Obs	204
25%	139.84	83.2	Sum of wgt.	204
50%	194.84		Mean	202.6036
		Largest	Std. dev.	86.02629
75%	241.725	423.28		
90%	292.88	431.43	Variance	7400.522
95%	332.13	583.12	Skewness	2.297989
99%	431.43	792.54	Kurtosis	14.396

```
19 .
20 . * Listando os 5 melhores e piores municípios em relação ao IDHM *
21 .
22 . sort idhm

23 . list mun idhm in 1/5 // Os 5 municípios com os piores IDHM
```

	mun	idhm
1.	Inhapi (AL)	.281
2.	Senador Rui Palmeira (AL)	.299
3.	Olivença (AL)	.301
4.	Canapi (AL)	.306
5.	Branquinha (AL)	.311

```
24 . list mun idhm in -5/l  // Os 5 municípios com melhores IDHM
```

	mun	idhm
200.	**Marechal Deodoro (AL)**	**.642**
201.	**Rio Largo (AL)**	**.643**
202.	**Arapiraca (AL)**	**.649**
203.	**Satuba (AL)**	**.66**
204.	**Maceió (AL)**	**.721**

```
25 .
26 .
27 . * a taxa de analfabetismo e a renda possuem uma relacao nao linear *
28 . * criando a variavel renda per capita ao quadrado *
29 .
30 . gen renda2 = renda^2

31 .
32 .
33 . ***** estimando através do modelo pooled *****
34 . * modelo MQO *
35 .
36 . reg analf gini dens idhm ensino_medio renda renda2
```

Source	SS	df	MS		
Model	**13800.8507**	**6**	**2300.14179**	Number of obs	= **204**
Residual	**2046.8299**	**197**	**10.3899995**	F(6, 197)	= **221.38**
				Prob > F	= **0.0000**
Total	**15847.6806**	**203**	**78.0673923**	R-squared	= **0.8708**
				Adj R-squared	= **0.8669**
				Root MSE	= **3.2234**

analf	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
gini	**12.0355**	**3.826212**	**3.15**	**0.002**	**4.489903**	**19.58109**
dens	**-.2603342**	**.0415985**	**-6.26**	**0.000**	**-.3423697**	**-.1782986**
idhm	**-37.09789**	**8.842143**	**-4.20**	**0.000**	**-54.53529**	**-19.66048**
ensino_medio	**-.5749367**	**.1147054**	**-5.01**	**0.000**	**-.8011448**	**-.3487286**
renda	**-.0608339**	**.0123958**	**-4.91**	**0.000**	**-.0852794**	**-.0363883**
renda2	**.0000491**	**.0000142**	**3.45**	**0.001**	**.0000211**	**.0000772**
_cons	**80.70733**	**5.31218**	**15.19**	**0.000**	**70.23129**	**91.18337**

```
37 .
38 . * Teste de Normalidade dos resíduos *
39 . predict uhat1, resid

40 . histogram uhat1, normal name("grafico1")
   (bin=14, start=-9.5216789, width=1.2068449)

41 .
42 . * Análise gráfica dos resíduos *
43 . gen lnanalf=ln(analf)

44 . scatter uhat1 lnanalf, yline(0) name("grafico2")

45 .
46 . * criando uma variavel logaritmica *
47 .
```

```
48 . gen lngini=ln(gini)

49 .
50 . * testando o modelo com a nova variavel *
51 .
52 . reg analf lngini dens idhm ensino_medio renda renda2
```

Source	SS	df	MS		
Model	13788.7213	6	2298.12021	Number of obs	204
				F(6, 197)	219.88
Residual	2058.95936	197	10.4515704	Prob > F	0.0000
				R-squared	0.8701
				Adj R-squared	0.8661
Total	15847.6806	203	78.0673923	Root MSE	3.2329

analf	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lngini	6.405162	2.174607	2.95	0.004	2.116665	10.69366
dens	-.2596863	.0417197	-6.22	0.000	-.3419609	-.1774118
idhm	-37.42798	8.866073	-4.22	0.000	-54.91258	-19.94338
ensino_medio	-.5762234	.1150402	-5.01	0.000	-.8030917	-.349355
renda	-.0606305	.0124328	-4.88	0.000	-.0851489	-.0361121
renda2	.0000493	.0000143	3.45	0.001	.0000211	.0000774
_cons	91.29759	4.676602	19.52	0.000	82.07496	100.5202

```
53 .
54 .
55 . ***** estimando através do modelo de efeitos fixos *****
56 .
57 . xtreg analf gini dens idhm ensino_medio renda renda2, fe
```

```
Fixed-effects (within) regression               Number of obs      =       204
Group variable: cod                             Number of groups   =       102

R-squared:                                      Obs per group:
     Within  = 0.9573                                         min =         2
     Between = 0.8148                                         avg =       2.0
     Overall = 0.7558                                         max =         2

                                                F(6,96)            =    358.38
corr(u_i, Xb) = 0.3427                          Prob > F           =    0.0000
```

analf	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
gini	6.404519	4.256634	1.50	0.136	-2.044832	14.85387
dens	-.06559	.0540798	-1.21	0.228	-.1729375	.0417576
idhm	-66.72144	7.252123	-9.20	0.000	-81.11679	-52.32609
ensino_medio	.003611	.1057375	0.03	0.973	-.2062763	.2134982
renda	-.0080508	.0125132	-0.64	0.522	-.0328893	.0167877
renda2	.0000152	.0000134	1.14	0.257	-.0000113	.0000418
_cons	72.93807	5.26794	13.85	0.000	62.48129	83.39485
sigma_u	4.4928739					
sigma_e	1.681971					
rho	.87707871	(fraction of variance due to u_i)				

```
F test that all u_i=0: F(101, 96) = 6.21                     Prob > F = 0.0000

58 .
```

```
59 .
60 . ***** estimando através do modelo de efeitos aleatórios *****
61 .
62 . xtreg analf gini dens idhm ensino_medio renda renda2, re

Random-effects GLS regression                   Number of obs     =       204
Group variable: cod                             Number of groups  =       102

R-squared:                                      Obs per group:
     Within  = 0.9408                                         min =         2
     Between = 0.8508                                         avg =       2.0
     Overall = 0.8640                                         max =         2

                                                Wald chi2(6)      =   1908.39
corr(u_i, X) = 0 (assumed)                      Prob > chi2       =    0.0000

------------------------------------------------------------------------------
       analf | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        gini |   12.71525   3.734861     3.40   0.001      5.39506    20.03545
        dens |  -.2457219   .0424371    -5.79   0.000     -.3288972   -.1625467
        idhm |  -51.82191   7.261037    -7.14   0.000    -66.05328   -37.59054
ensino_medio |   -.304579   .1002238    -3.04   0.002     -.501014   -.1081441
       renda |   -.045826   .0113329    -4.04   0.000     -.0680381   -.0236139
      renda2 |   .0000345   .0000128     2.70   0.007      9.45e-06    .0000596
       _cons |   81.15234   4.789269    16.94   0.000      71.76555    90.53913
-------------+----------------------------------------------------------------
     sigma_u |  2.3494466
     sigma_e |   1.681971
         rho |  .66115084   (fraction of variance due to u_i)
------------------------------------------------------------------------------

63 .
64 .
65 .
66 . ***** utilizando o teste de Hausman para escolher qual modelo é mais adequado *****
67 .
68 . qui xtreg analf gini dens idhm ensino_medio renda renda2, fe

69 . estimates store fe

70 .
71 . qui xtreg analf gini dens idhm ensino_medio renda renda2, re

72 . estimates store re

73 .
74 . hausman fe re

Note: the rank of the differenced variance matrix (5) does not equal the number of coefficients being tested (6); be
        you expect, or there may be problems computing the test.  Examine the output of your estimators for anything
        possibly consider scaling your variables so that the coefficients are on a similar scale.

                 ---- Coefficients ----
                 (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
                  fe           re         Difference       Std. err.
-------------+----------------------------------------------------------------
        gini |   6.404519    12.71525      -6.310734        2.041995
        dens |    -.06559   -.2457219       .1801319        .0335219
        idhm |  -66.72144   -51.82191      -14.89953              .
ensino_medio |    .003611    -.304579         .30819        .0336989
       renda |  -.0080508    -.045826       .0377752        .0053052
      renda2 |   .0000152    .0000345      -.0000193        3.93e-06
------------------------------------------------------------------------------
                           b = Consistent under H0 and Ha; obtained from xtreg.
            B = Inconsistent under Ha, efficient under H0; obtained from xtreg.

Test of H0: Difference in coefficients not systematic
```

```
    chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
            =  95.70
Prob > chi2 = 0.0000
(V_b-V_B is not positive definite)

75 .
76 .
77 . ***** utilizando o teste de Breush-Pagan *****
78 . * reestima-se o modelo de EA *
79 . xtreg analf gini dens idhm ensino_medio renda renda2, re

Random-effects GLS regression                   Number of obs     =       204
Group variable: cod                             Number of groups  =       102

R-squared:                                      Obs per group:
     Within  = 0.9408                                         min =         2
     Between = 0.8508                                         avg =       2.0
     Overall = 0.8640                                         max =         2

                                                Wald chi2(6)      =   1908.39
corr(u_i, X) = 0 (assumed)                      Prob > chi2       =    0.0000
```

analf	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
gini	12.71525	3.734861	3.40	0.001	5.39506	20.03545
dens	-.2457219	.0424371	-5.79	0.000	-.3288972	-.1625467
idhm	-51.82191	7.261037	-7.14	0.000	-66.05328	-37.59054
ensino_medio	-.304579	.1002238	-3.04	0.002	-.501014	-.1081441
renda	-.045826	.0113329	-4.04	0.000	-.0680381	-.0236139
renda2	.0000345	.0000128	2.70	0.007	9.45e-06	.0000596
_cons	81.15234	4.789269	16.94	0.000	71.76555	90.53913
sigma_u	2.3494466					
sigma_e	1.681971					
rho	.66115084	(fraction of variance due to u_i)				

```
80 . xttest0

Breusch and Pagan Lagrangian multiplier test for random effects

        analf[cod,t] = Xb + u[cod] + e[cod,t]

        Estimated results:
                         |       Var     SD = sqrt(Var)
                ---------+-----------------------------
                   analf |   78.06739       8.835575
                       e |   2.829027       1.681971
                       u |   5.519899       2.349447

        Test: Var(u) = 0
                                 chibar2(01) =    21.16
                              Prob > chibar2 =   0.0000

81 .
82 .
83 . ***** análise gráfica com duas variáveis ****
84 . twoway fpfitci analf renda || scatter analf renda, name("grafico3")

85 . twoway (scatter analf renda, mcolor(black)) (scatter analf renda if mun=="Maceió (AL)", mlabel(mun)),graphregion(c
   > ion(color(white)) name("grafico4")
```

```
86 .
87 .
88 . **** Criando gráfico box-plot para as variáveis *********
89 . graph box analf ensino_medio, name("grafico5")

90 . gen lnrenda=ln(renda)

91 . graph box analf gini dens idhm ensino_medio lnrenda, name("grafico6")

92 .
   end of do-file

93 .
```