Statistics/Data analysis

```
      name:  <unnamed>
       log:  A:\_maestria_unibo_(operacional)\4_econometrics_1\4_problem_sets\2_ps2\3_
> log\log.smcl
  log type:  smcl
 opened on:  27 Oct 2021, 19:59:47

1 .
2 . //*#########################################################################
3 . //* A.
4 . //*#########################################################################
5 .
6 . use 1_data\dataset_1.dta

7 .
8 . //*=========================================================================
9 . //* A.1.
10. //*=========================================================================
11. //*----- Calculate the mean value for each TV over all sample.
12.
13.
14. //*----- Calculate mean values of each target variable (TV).
15.
16. bysort state_code: egen mean_etr = mean(expos_to_robots)

17. bysort state_code: egen mean_emp = mean(d_emppriv_1990_2011)

18.
19. //*----- Remove duplicated data to display min and max values.
20.
21. quietly by state_code:  gen dup = cond(_N==1,0,_n)

22. drop if dup > 1
  (674 observations deleted)

23.
24. //*----- Identify the states that have the min and max values for each TV.
25.
26. sort mean_etr

27. gen min_etr_state = 1 if _n == 1
  (47 missing values generated)

28. gen max_etr_state = 1 if _n == _N
  (47 missing values generated)

29. table state_code (min_etr_state), statistic (mean mean_etr) nototals

  ----------------------------------
                  | min_etr_state
                  |             1
  ----------------+-----------------
  state_l         |
    NV            |     .5719783
  ----------------------------------

30. table state_code (max_etr_state), statistic (mean mean_etr) nototals

  ----------------------------------
                  | max_etr_state
                  |             1
  ----------------+-----------------
  state_l         |
    MI            |     5.246422
  ----------------------------------
```

```
31.
32. sort mean_emp

33. gen min_emp_state = 1 if _n == 1
  (47 missing values generated)

34. gen max_emp_state = 1 if _n == _N
  (47 missing values generated)

35. table state_code (min_emp_state), statistic (mean mean_emp) nototals
```

	min_emp_state
	1
state_l	
NC	**-4.594934**

```
36. table state_code (max_emp_state), statistic (mean mean_emp) nototals
```

	max_emp_state
	1
state_l	
ND	**7.900403**

```
37.
38. //*----- Calculate the mean value for each TV over state means.
39.
40. summarize mean_etr
```

Variable	Obs	Mean	Std. dev.	Min	Max
mean_etr	**48**	**1.852853**	**.9572189**	**.5719783**	**5.246422**

```
41. summarize mean_emp
```

Variable	Obs	Mean	Std. dev.	Min	Max
mean_emp	**48**	**.3809222**	**2.482054**	**-4.594934**	**7.900403**

```
42.
43. //*----- Two-way catterplot with regression line.
44.
45. clear all

46. use 1_data\dataset_1.dta

47.
48. graph twoway (lfitci d_empprivˍ1990ˍ2011 expos_to_robots, xtitle("Exposure to robots
  > ")) (scatter d_empprivˍ1990ˍ2011 expos_to_robots, msize(tiny) legend(size(small) lab
  > el(3 "Change in the share of private employment over total population. 1990-2011."))
  > )

49.
50. graph save "scatter_line.gph", replace
  file scatter_line.gph saved
```

```
51. graph save 4_graphs\scatter_line.gph, replace
  file 4_graphs\scatter_line.gph saved

52. graph export 4_graphs\scatter_line.png, as(png) replace
  file 4_graphs\scatter_line.png saved as PNG format

53. graph close

54.
55. //*#############################################################################
56. //* B.
57. //*#############################################################################
58.
59. clear all

60. use 1_data\dataset_1.dta

61.
62. //*=============================================================================
63. //* B.2. Simple linear regression.
64. //*=============================================================================
65.
66. reg d_emppriv_1990_2011 expos_to_robots, rob

  Linear regression                               Number of obs     =        722
                                                  F(1, 720)         =      99.20
                                                  Prob > F          =     0.0000
                                                  R-squared         =     0.1421
                                                  Root MSE          =      2.955

------------------------------------------------------------------------------
             |               Robust
d_emppriv_~2011 |  Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
expos_to_robots |   -1.09071   .1095083    -9.96   0.000    -1.305704   -.8757165
          _cons |   2.774654   .2344016    11.84   0.000     2.314461    3.234846
------------------------------------------------------------------------------

67. outreg2 using 5_tables\reg1, tex replace
  5_tables\reg1.tex
  dir : seeout

68.
69. //*=============================================================================
70. //* B.3. Regression with additional controls.
71. //*=============================================================================
72.
73. reg d_emppriv_1990_2011 expos_to_robots ipums_logpop_1990 ipums_female_1990 ipums_ab
  > ove65_1990 ipums_highschool_1990 ipums_somecollege_1990 ipums_college_1990 ipums_mas
  > ters_1990 ipums_white_1990 ipums_black_1990 ipums_hispanic_1990 ipums_asian_1990, ro
  > b
  note: ipums_college_1990 omitted because of collinearity.
  note: ipums_hispanic_1990 omitted because of collinearity.

  Linear regression                               Number of obs     =        722
                                                  F(10, 711)        =      46.76
                                                  Prob > F          =     0.0000
                                                  R-squared         =     0.3847
                                                  Root MSE          =     2.5183

------------------------------------------------------------------------------
             |               Robust
  d_emppriv_1990_2011 |  Coefficient  std. err.      t    P>|t|     [95% conf. interva
> l]
-------------+----------------------------------------------------------------
      expos_to_robots |   -.555736   .0887673    -6.26   0.000    -.7300135   -.38145
> 85
   ipums_logpop_1990 |  -.4474018   .0894417    -5.00   0.000    -.6230033   -.27180
> 03
   ipums_female_1990 |   .6339813   15.73012     0.04   0.968    -30.24906    31.517
```

```
> 02
    ipums_above65_1990 |   11.57593   5.439096     2.13   0.034     .8973153    22.254
> 54
 ipums_highschool_1990 |  -23.29328   6.426127    -3.62   0.000    -35.90973   -10.676
> 83
ipums_somecollege_1990 |  -12.70352    8.72179    -1.46   0.146    -29.82706    4.4200
> 27
    ipums_college_1990 |          0  (omitted)
    ipums_masters_1990 |  -76.55848   15.78116    -4.85   0.000    -107.5417   -45.575
> 23
      ipums_white_1990 |  -1.879022   .6873117    -2.73   0.006    -3.228426   -.52961
> 92
      ipums_black_1990 |  -7.003238   .9646186    -7.26   0.000    -8.897079   -5.1093
> 96
   ipums_hispanic_1990 |          0  (omitted)
      ipums_asian_1990 |  -28.80399   8.598956    -3.35   0.001    -45.68638   -11.921
> 61
                 _cons |   27.46541   11.39099     2.41   0.016     5.101417     49.82
> 94
----------------------+
  --
```

```
74. outreg2 using 5_tables\reg2, tex replace
  5_tables\reg2.tex
  dir : seeout

75.
76. //*----- Dropping some variables because of multicollinearity.
77.
78. reg d_emppriv_1990_2011 expos_to_robots ipums_logpop_1990 ipums_female_1990 ipums_ab
  > ove65_1990 ipums_somecollege_1990 ipums_college_1990 ipums_masters_1990 ipums_black_
  > 1990 ipums_hispanic_1990 ipums_asian_1990, rob

Linear regression                               Number of obs     =        722
                                                F(10, 711)        =      46.76
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3847
                                                Root MSE          =     2.5183

------------------------+
  --
                        |               Robust
  d_emppriv_1990_2011   | Coefficient  std. err.      t    P>|t|     [95% conf. interva
> l]
------------------------+
  --
     expos_to_robots    |   -.555736   .0887673    -6.26   0.000    -.7300135   -.38145
> 85
   ipums_logpop_1990    |  -.4474018   .0894417    -5.00   0.000    -.6230033   -.27180
> 03
   ipums_female_1990    |   .6339831   15.73012     0.04   0.968    -30.24906    31.517
> 02
   ipums_above65_1990   |   11.57593   5.439096     2.13   0.034     .8973147    22.254
> 54
ipums_somecollege_1990  |   10.58976    3.28857     3.22   0.001     4.133291    17.046
> 23
   ipums_college_1990   |   23.29329   6.426127     3.62   0.000     10.67683    35.909
> 74
   ipums_masters_1990   |  -53.26521    11.6925    -4.56   0.000    -76.22117   -30.309
> 24
     ipums_black_1990   |  -5.124215   1.046752    -4.90   0.000    -7.179309   -3.0691
> 22
  ipums_hispanic_1990   |   1.879022   .6873117     2.73   0.006     .5296187    3.2284
> 25
     ipums_asian_1990   |  -26.92497   8.413793    -3.20   0.001    -43.44382   -10.406
> 12
                 _cons  |   2.293104   7.817483     0.29   0.769    -13.05501    17.641
> 22
------------------------+
  --
```

```
79. outreg2 using 5_tables\reg2, tex replace
  5_tables\reg2.tex
  dir : seeout

80.
81. //*============================================================================
82. //* B.4. Regression with the square of expos_to_robots.
83. //*============================================================================
84.
85. gen expos_2 = expos_to_robots*expos_to_robots

86. reg d_emppriv_1990_2011 expos_to_robots expos_2 ipums_logpop_1990 ipums_female_1990
  > ipums_above65_1990 ipums_somecollege_1990 ipums_college_1990 ipums_masters_1990 ipum
  > s_black_1990 ipums_hispanic_1990 ipums_asian_1990, rob

  Linear regression                               Number of obs     =        722
                                                  F(11, 710)        =      47.47
                                                  Prob > F          =     0.0000
                                                  R-squared         =     0.3934
                                                  Root MSE          =     2.5022

------------------------------------------------------------------------------
             |               Robust
d_emppriv_1990_2011 | Coefficient  std. err.      t    P>|t|     [95% conf. interva
> l]
-------------+----------------------------------------------------------------
     expos_to_robots |   -1.32494   .2462026    -5.38   0.000    -1.808312   -.84156
> 78
            expos_2 |   .1055034   .0272236     3.88   0.000     .0520549    .15895
> 18
   ipums_logpop_1990 |   -.404733   .0928466    -4.36   0.000    -.5870197   -.22244
> 63
   ipums_female_1990 |   4.906006   15.72564     0.31   0.755    -25.96831    35.780
> 32
  ipums_above65_1990 |   10.56819   5.561721     1.90   0.058    -.3512001    21.487
> 57
ipums_somecollege_1990 |   8.748911   3.285391     2.66   0.008     2.298667    15.199
> 16
  ipums_college_1990 |   20.84298   6.385557     3.26   0.001     8.306145    33.379
> 81
  ipums_masters_1990 |  -50.07539   11.56498    -4.33   0.000    -72.78104   -27.369
> 74
   ipums_black_1990 |  -5.610725   1.052319    -5.33   0.000    -7.676753   -3.5446
> 97
 ipums_hispanic_1990 |   1.222744   .7225081     1.69   0.091    -.1957636    2.6412
> 52
   ipums_asian_1990 |  -26.26183   8.474227    -3.10   0.002    -42.89937   -9.6242
> 86
               _cons |   1.340348   7.813079     0.17   0.864    -13.99915    16.679
> 85
------------------------------------------------------------------------------

87. outreg2 using 5_tables\reg3, tex replace
  5_tables\reg3.tex
  dir : seeout

88.
```

```
89. //*=============================================================================
90. //* B.5. Testing the marginal effect of expos_to_robots (and its cuadratic term).
91. //*=============================================================================
92. //*----- Evaluating the marginal effect of exposure to robots at the mean of expos_t
  > o_robots.
93.
94. summarize expos_to_robots
```

Variable	Obs	Mean	Std. dev.	Min	Max
expos_to_r~s	**722**	**1.800013**	**1.101769**	**.4188264**	**10.26309**

```
95. scalar mean_etr = r(mean)

96. di mean_etr
  1.800013

97. lincom _b[expos_to_robots] + (2 * _b[expos_2] * mean_etr)

   ( 1)  expos_to_robots + 3.600026*expos_2 = 0
```

d_emppr~2011	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
(1)	**-.9451252**	**.1541927**	**-6.13**	**0.000**	**-1.247853**	**-.6423971**

```
98.
99. //*=============================================================================
100 //* B.6. Heteroskedasticity tests.
101 //*----- Runnign full regression without rob.
102
103 reg d_emppriv_1990_2011 expos_to_robots expos_2 ipums_logpop_1990 ipums_female_1990
  > ipums_above65_1990 ipums_somecollege_1990 ipums_college_1990 ipums_masters_1990 ipum
  > s_black_1990 ipums_hispanic_1990 ipums_asian_1990
```

Source	SS	df	MS		
Model	**2882.94304**	**11**	**262.085731**	Number of obs	= **722**
				F(11, 710)	= **41.86**
Residual	**4445.29832**	**710**	**6.26098355**	Prob > F	= **0.0000**
				R-squared	= **0.3934**
				Adj R-squared	= **0.3840**
Total	**7328.24136**	**721**	**10.1639963**	Root MSE	= **2.5022**

d_emppriv_1990_2011	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
expos_to_robots	**-1.32494**	**.2619301**	**-5.06**	**0.000**	**-1.83919**	**-.81068**
expos_2	**.1055034**	**.0330962**	**3.19**	**0.001**	**.0405253**	**.17048**
ipums_logpop_1990	**-.404733**	**.0852326**	**-4.75**	**0.000**	**-.5720711**	**-.23739**
ipums_female_1990	**4.906006**	**15.32995**	**0.32**	**0.749**	**-25.19145**	**35.003**
ipums_above65_1990	**10.56819**	**4.401876**	**2.40**	**0.017**	**1.925936**	**19.210**
ipums_somecollege_1990	**8.748911**	**3.116931**	**2.81**	**0.005**	**2.629407**	**14.868**
ipums_college_1990	**20.84298**	**5.934436**	**3.51**	**0.000**	**9.191836**	**32.494**
ipums_masters_1990	**-50.07539**	**12.29295**	**-4.07**	**0.000**	**-74.21027**	**-25.94**
ipums_black_1990	**-5.610725**	**1.126267**	**-4.98**	**0.000**	**-7.821937**	**-3.3995**
ipums_hispanic_1990	**1.222744**	**.9225356**	**1.33**	**0.185**	**-.5884797**	**3.0339**
ipums_asian_1990	**-26.26183**	**10.24617**	**-2.56**	**0.011**	**-46.37824**	**-6.1454**
_cons	**1.340348**	**7.462319**	**0.18**	**0.858**	**-13.3105**	**15.99**

```
> 12
```

```
104 rvfplot, yline(0)

105 estat imtest, white

  White's test
  H0: Homoskedasticity
  Ha: Unrestricted heteroskedasticity

     chi2(76) = 168.18
  Prob > chi2 = 0.0000

  Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	**168.18**	**76**	**0.0000**
Skewness	**17.13**	**11**	**0.1042**
Kurtosis	**9.52**	**1**	**0.0020**
Total	**194.83**	**88**	**0.0000**

```
106
107 //*#########################################################################
108 //* C.
109 //*#########################################################################
110
111 clear all

112 use 1_data\dataset_2.dta

113 summarize
```

Variable	Obs	Mean	Std. dev.	Min	Max
czone	**23,770**	**21937.98**	**11352.08**	**100**	**39400**
d_yrwag~2011	**23,770**	**.0813208**	**.6081289**	**-4.947901**	**4.888123**
state_code	**23,770**	**24.8496**	**13.81532**	**1**	**48**
expos_to_r~s	**23,770**	**1.842783**	**1.157471**	**.4188264**	**10.26309**
education	**23,770**	**2.812873**	**1.318574**	**1**	**5**
race	**23,770**	**3.287042**	**1.424595**	**1**	**5**
female	**23,770**	**.4982751**	**.5000075**	**0**	**1**

```
114
115 //*=========================================================================
116 //* C.1. Creating a dummy variable for whites.
117 //*=========================================================================
118
119 codebook race
```

race **(unlabeled)**

Type: Numeric (**long**)
Label: **race_num**

Range: [**1,5**] Units: **1**
Unique values: **5** Missing .: **0**/**23,770**

```
            Tabulation: Freq.   Numeric  Label
                        3,347         1  asian
                        4,426         2  black
                        5,274         3  hispanic
                        3,503         4  other
                        7,220         5  white

120 gen white = 0 if race != 5
  (7,220 missing values generated)

121 replace white = 1 if race == 5
  (7,220 real changes made)

122 tab race white
```

race	white 0	1	Total
asian	3,347	0	3,347
black	4,426	0	4,426
hispanic	5,274	0	5,274
other	3,503	0	3,503
white	0	7,220	7,220
Total	16,550	7,220	23,770

```
123
124 //*=================================================================================
125 //* C.2. Regression.
126 //*=================================================================================
127
128 reg d_yrwage_ln_1990_2011 c.expos_to_robots##i.female white##i.education, rob

  Linear regression                               Number of obs     =     23,770
                                                  F(12, 23757)      =     189.29
                                                  Prob > F          =     0.0000
                                                  R-squared         =     0.0368
                                                  Root MSE          =     .59698

------------------------------------------------------------------------------
             |               Robust
d_yrwage_ln_1990_2011 | Coefficient  std. err.      t    P>|t|     [95% conf. inter
> val]
-------------+----------------------------------------------------------------
        expos_to_robots |  -.0452585   .0044846   -10.09   0.000    -.0540486   -.036
> 4685
              1.female |   .1428625   .0140142    10.19   0.000     .1153937    .170
> 3312
                       |
female#c.expos_to_robots |
                     1 |   .0078448   .0063929     1.23   0.220    -.0046856    .020
> 3752
                       |
               1.white |  -.0301436   .0146348    -2.06   0.039    -.0588287   -.001
> 4585
                       |
             education |
          high school |  -.0310826   .0167027    -1.86   0.063     -.063821    .001
> 6558
        some college |  -.0251477   .0171907    -1.46   0.144    -.0588425    .008
> 5471
             college |   .1015417   .0191859     5.29   0.000     .0639361    .139
> 1472
              master |   .1624391   .0202855     8.01   0.000     .1226782    .202
> 2001
                       |
      white#education |
       1#high school |   .0469742   .0180229     2.61   0.009      .011648    .082
> 3003
      1#some college |    .104628   .0184406     5.67   0.000     .0684833    .140
```

> 7727

1#college	.0593417	.0204027	2.91	0.004	.0193511	.099
> 3323						
1#master	.0211344	.0217293	0.97	0.331	-.0214564	.063
> 7251						
_cons	.0541556	.0160536	3.37	0.001	.0226895	.085
> 6218						

```
------------------------------------------------------------------------------
```

```
129 outreg2 using 5_tables\reg4, tex replace
  5_tables\reg4.tex
  dir : seeout

130
131 //*============================================================================
132 //* C.3. Chow test for testing differences between whites and non-whites.
133 //*============================================================================
134
135 //* Not included.
136
137 //*============================================================================
138 //* C.4 Test of the difference in marginal effect of robots on men and women.
139 //*============================================================================
140
141 test 1.female#c.expos_to_robots

   ( 1)  1.female#c.expos_to_robots = 0

         F(  1, 23757) =    1.51
              Prob > F =    0.2198

142
143 //*============================================================================
144 //* C.5. Regression with dummy variables for education levels.
145 //*============================================================================
146
147 tabulate education, generate(edu)
```

education	Freq.	Percent	Cum.
less than high school	4,810	20.24	20.24
high school	5,643	23.74	43.98
some college	5,732	24.11	68.09
college	4,355	18.32	86.41
master	3,230	13.59	100.00
Total	23,770	100.00	

```
148 reg d_yrwage_ln_1990_2011 expos_to_robots female edu2 edu3 edu4 edu5
```

Source	SS	df	MS		
Model	315.098903	6	52.5164838	Number of obs	= 23,770
Residual	8475.1713	23,763	.356654097	F(6, 23763)	= 147.25
				Prob > F	= 0.0000
Total	8790.2702	23,769	.369820783	R-squared	= 0.0358
				Adj R-squared	= 0.0356
				Root MSE	= .59721

d_yrwage_l~2011	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
expos_to_robots	-.0414809	.0033469	-12.39	0.000	-.048041	-.0349208
female	.1572163	.0077475	20.29	0.000	.1420307	.1724019
edu2	-.0176851	.0117199	-1.51	0.131	-.0406569	.0052867
edu3	.0027235	.0116779	0.23	0.816	-.0201659	.0256129
edu4	.1202807	.0124924	9.63	0.000	.0957947	.1447667
edu5	.1674346	.0135864	12.32	0.000	.1408045	.1940648
_cons	.0381767	.0112288	3.40	0.001	.0161675	.0601859

```
149
150 //*============================================================================
151 //* C.6. Regression with interaction between robots and dummies for education.
152 //*============================================================================
153
154 reg d_yrwage_ln_1990_2011 c.expos_to_robots##i.education female race, rob

Linear regression                               Number of obs     =     23,770
                                                F(11, 23758)      =      89.82
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0370
                                                Root MSE          =     .59693

------------------------------------------------------------------------------
             |               Robust
d_yrwage_ln_1990_2011 | Coefficient  std. err.      t    P>|t|     [95% conf. in
> terval]
-------------+----------------------------------------------------------------
      expos_to_robots |  -.0557862   .0083043    -6.72   0.000    -.0720631   -.
> 0395093
                      |
           education |
          high school |  -.0100199   .0223958    -0.45   0.655    -.0539171    .
> 0338772
        some college |  -.0223304   .0226929    -0.98   0.325    -.0668099    .
> 0221491
             college |   .0529709   .0248537     2.13   0.033     .0042561    .
> 1016858
              master |   .0995622   .0233687     4.26   0.000     .0537581    .
> 1453663
                      |
education#c.expos_to_robots |
          high school |  -.0037766   .0103959    -0.36   0.716    -.0241532
>   .0166
        some college |   .0140973   .0102976     1.37   0.171    -.0060867    .
> 0342813
             college |   .0367525   .0115503     3.18   0.001     .0141132    .
> 0593917
              master |   .0363174   .0108219     3.36   0.001     .0151058    .
> 0575291
                      |
              female |   .1573782   .0077434    20.32   0.000     .1422006    .
> 1725558
                race |   .0040804   .0026976     1.51   0.130    -.0012071    .
> 0093679
               _cons |   .0505455   .0214288     2.36   0.018     .0085436    .
> 0925473
------------------------------------------------------------------------------

155 outreg2 using 5_tables\reg5, tex replace
  5_tables\reg5.tex
  dir : seeout

156
157 //*############################################################################
158 //* n. Close log.
159 //*############################################################################
160
161 log close
       name:  <unnamed>
        log:  A:\_maestria_unibo_(operacional)\4_econometrics_1\4_problem_sets\2_ps2\3_
> log\log.smcl
   log type:  smcl
  closed on:  27 Oct 2021, 19:59:53
-------------------------------------------------------------------------------
```