Reihaneh Moghisi Jean Marc Kangah Yagaize anyanwu

Econ8740
Assignment 2
September 28 2022

1. [21 Points] The data contain 935 observations on 17 variables. Use the describe command to get definition of all 17 variables.
Obtain summary statistics for all variables in the data and note the number of missing values for educ, sibs, meduc, and feduc.

```
1 . describe

Contains data from /Users/reihaneh/Desktop/Georgia state University/Fall2022/ECON 8740 e
> onometrics/Assignments/2/WAGE2.DTA
  obs:           935
 vars:            17                          14 Apr 1999 13:41
 size:        20,570
```

variable name	storage type	display format	value label	variable label
wage	int	%9.0g		**monthly earnings**
hours	byte	%9.0g		**average weekly hours**
IQ	int	%9.0g		**IQ score**
KWW	byte	%9.0g		**knowledge of world work score**
educ	byte	%9.0g		**years of education**
exper	byte	%9.0g		**years of work experience**
tenure	byte	%9.0g		**years with current employer**
age	byte	%9.0g		**age in years**
married	byte	%9.0g		**=1 if married**
black	byte	%9.0g		**=1 if black**
south	byte	%9.0g		**=1 if live in south**
urban	byte	%9.0g		**=1 if live in SMSA**
sibs	byte	%9.0g		**number of siblings**
brthord	byte	%9.0g		**birth order**
meduc	byte	%9.0g		**mother's education**
feduc	byte	%9.0g		**father's education**
lwage	float	%9.0g		**natural log of wage**

```
Sorted by:
```

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
wage	935	957.9455	404.3608	115	3078
hours	935	43.92941	7.224256	20	80
IQ	935	101.2824	15.05264	50	145
KWW	935	35.74439	7.638788	12	56
educ	935	13.46845	2.196654	9	18
exper	935	11.56364	4.374586	1	23
tenure	935	7.234225	5.075206	0	22
age	935	33.08021	3.107803	28	38
married	935	.8930481	.3092174	0	1
black	935	.1283422	.3346495	0	1
south	935	.3411765	.4743582	0	1
urban	935	.7176471	.4503851	0	1
sibs	935	2.941176	2.306254	0	14
brthord	852	2.276995	1.595613	1	10
meduc	857	10.68261	2.849756	0	18
feduc	741	10.21727	3.3007	0	18
lwage	935	6.779004	.4211439	4.744932	8.032035

Note the number of missing values for educ, sibs, meduc, feduc

```
. count if missing(educ)
  0

. count if missing(sibs)
  0

. count if missing(meduc)
  78

. count if missing(feduc)
  194
```

A. Obtain basic summary statistics for educ, sibs, meduc, and feduc for non-missing observations of meduc and feduc. (Hint: You should get 722 observations for each of the 4 variables.)

```
2 . sum educ sibs meduc feduc if meduc ~=. & feduc ~=.
```

Variable	Obs	Mean	Std. Dev.	Min	Max
educ	722	13.66343	2.236755	9	18
sibs	722	2.858726	2.250471	0	14
meduc	722	10.80609	2.828636	0	18
feduc	722	10.25485	3.305757	0	18

B- Estimate a linear regression of educ on sibs, meduc, and feduc. Write down the predicted equation. How much of the variation in educ is explained by sibs, meduc and feduc?
Use the results to answer parts 1C through 1I below.

```
. regress educ sibs meduc feduc
```

Source	SS	df	MS	Number of obs	=	722
				F(3, 718)	=	65.20
Model	772.281437	3	257.427146	Prob > F	=	0.0000
Residual	2834.93324	718	3.94837499	R-squared	=	0.2141
				Adj R-squared	=	0.2108
Total	3607.21468	721	5.00307168	Root MSE	=	1.9871

educ	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sibs	-.0936359	.0344713	-2.72	0.007	-.1613124	-.0259594
meduc	.1307872	.032689	4.00	0.000	.0666098	.1949646
feduc	.2100041	.0274748	7.64	0.000	.1560635	.2639447
_cons	10.36426	.3585001	28.91	0.000	9.660422	11.06809

The predicted equation is:

Educ = 10.36426 - 0.0936359*Sibs + 0.1307872*meduc +0.2100041*feduc

Based on R-square, 21.41% of variation in educ is explained by independent variables including sibs, meduc and feduc

C- Does sibs have the expected effect? Explain
There is a negative relation between the number of siblings and the number of years of education in a family. According to the results of the predicted equation, one more sibling reduced the number of years of education by 0.0936359. This can be interpreted as when the number of children increase in a family, ceteris paribus, the parents will have less time to spend for each child and less money to pay school fees.

Holding meduc and feduc fixed, by how much does sibs have to increase to reduce predicted years of education by one year? (A non-integer answer is acceptable here).
$\Delta Educ = - 0.0936359*\Delta Sibs$
$1 = - 0.0936359*\Delta Sibs$
$\Delta Sibs = 1/- 0.0936359$
$\Delta Sibs = -10.67.$
To reduce the predicted years of education by one year, holding everything else constant,Sibs has to increase by 10.67.

D. Discuss the interpretation of the coefficient on feduc

One year increase in the level of father's education affects education by 0.2 if holding everything else constant.

E. Suppose that Man A has no siblings, and his mother and father each have 10 years of education. Man B has no siblings, and his mother and father each have 15 years of education.

What is the predicted difference in years of education between B and A?

Educ $_{ManA}$= 10.36426 - 0.0936359*(0) + 0.1307872*(10) +0.2100041*(10)
Educ $_{ManA}$-= **13.7721**
Educ $_{ManB}$ =10.36426 - 0.0936359*(0) + 0.1307872*(15) +0.2100041*(15)
Educ $_{ManB}$= **15.4761**
The predicted difference between ManA and ManB is: 15.4761- 13.7721 = 1.704.

F. Which of the factors are individually statistically significant at the 5% level of significance?

All the factors including sibs, meduc and feduc are statistically significant on educ
P>|t|
Sibs: 0.007
Meduc: 0.000
Feduc: 0.000

G. Construct the 99% confidence interval for each of the population coefficients associated with the explanatory variables: sibs, meduc, and feduc.

```
. regress educ sibs meduc feduc, level(99)
```

Source	SS	df	MS		
				Number of obs	= 722
				F(3, 718)	= 65.20
Model	772.281437	3	257.427146	Prob > F	= 0.0000
Residual	2834.93324	718	3.94837499	R-squared	= 0.2141
				Adj R-squared	= 0.2108
Total	3607.21468	721	5.00307168	Root MSE	= 1.9871

educ	Coef.	Std. Err.	t	P>|t|	[99% Conf.	Interval]
sibs	-.0936359	.0344713	-2.72	0.007	-.1826646	-.0046072
meduc	.1307872	.032689	4.00	0.000	.0463616	.2152128
feduc	.2100041	.0274748	7.64	0.000	.139045	.2809631
_cons	10.36426	.3585001	28.91	0.000	9.43836	11.29015

[99% Conf. Interval]
Sibs: (-0.1826646 , -0.0046072)
Meduc: (0.0463616, 0.2152128)
Feduc: (0.139045, 0.2809631)

H. Test the joint significance of sibs, meduc and feduc.

For the regression equation: Educ = β_0+ β_1*Sibs + β_2*meduc +β_3*feduc
null hypothesis is: H_0: β1=β2 = β3 = 0
Compute the F-statistic: Fq, N-k = (RSSR – RSSUR) / q

```
. test sibs meduc feduc

 ( 1)  sibs = 0
 ( 2)  meduc = 0
 ( 3)  feduc = 0

       F(  3,   718) =   65.20
            Prob > F =    0.0000
```

I. Test the joint significance of meduc and feduc.

```
. test meduc feduc

 ( 1)  meduc = 0
 ( 2)  feduc = 0

       F(  2,   718) =   77.11
            Prob > F =    0.0000
```

Question 2.

2 provide des and detail of data
des,detail

```
37 . * 2 provide des and detail of data
38 . des,detail

Contains data from /Users/reihaneh/Desktop/Georgia state University/Fall2022/ECON 8740
>  Stat and Econometrics/Assignments/2/data7aweb.dta
  obs:       426,792
 vars:            24                          30 Jul 2001 07:15
width:           115
 size:    49,081,080
-------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
cty1            int     %9.0g                 IFS Country Code 1
cty2            int     %9.0g                 IFS Country Code 2
year            int     %8.0g
pairid          int     %9.0g                 Unique Country-Pair Identifier
landl           byte    %9.0g                 # Landlocked 0/1/2
island          byte    %9.0g                 # Islands 0/1/2
border          byte    %8.0g                 Land Border Dummy
comlang         byte    %8.0g                 1 for Common Language
comcol          byte    %8.0g                 Dummy for Common Colonizer post 1945
comctry         byte    %8.0g                 Dummy for Same Nation/Perennial Colonies
colony          byte    %8.0g                 Dummy for pairs ever in Colonial
                                                Relationship
curcol          byte    %8.0g                 Dummy for pairs currently in Colonial
                                                Relationship
custrict        byte    %8.0g                 Strict Currency Union
cumed           byte    %8.0g                 Strict or inferred (from Transitivity)
                                                Currency Union
ltrade          float   %9.0g                 Log Value of Bilateral Trade in Real $
regional        byte    %8.0g                 RTA Dummy
lareap          float   %9.0g                 Log of Product of Land Areas

ldist           float   %9.0g                 Log of Distance
lrgdp           float   %9.0g                 Log of Product of Real GDPs
lrgdppc         float   %9.0g                 Log of Product of Real GDPs per capita
ctyname1        str34   %34s
rgdp1           float   %9.0g                 Real GDP
ctyname2        str34   %34s
rgdp2           float   %9.0g                 Real GDP
-------------------------------------------------------------------------------
Sorted by: cty1  cty2  year
```

* 2.1 define the set of variables called A

global A ltrade ldist lrgdp lrgdppc comlang border regional landl island lareap comcol curcol colony comctry

* 2.1 Provide descriptive statistics of variables in A by currencyunion

bysort custrict: sum $A

```
. * 2.1 define the set of variables called A
. global A ltrade ldist lrgdp lrgdppc comlang border regional landl island lareap comc
> ol curcol colony comctry

.
. * 2.1 Provide descriptive statistics by currencyunion
. bysort custrict: sum $A
```

-> custrict = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
ltrade	422,715	10.70738	3.684208	-16.11561	23.86674
ldist	306,720	8.150077	.8195516	3.684131	9.421514
lrgdp	216,389	47.89685	2.632033	36.87086	58.01698
lrgdppc	216,389	16.06051	1.436249	10.1211	20.89841
comlang	422,715	.146503	.3536102	0	1
border	422,715	.0200348	.1401193	0	1
regional	422,715	.0065269	.080525	0	1
landl	306,727	.2314892	.4546681	0	2
island	306,727	.3457798	.5433083	0	2
lareap	306,727	23.78475	3.609643	4.916325	33.03472
comcol	422,715	.0590634	.2357437	0	1
curcol	422,715	.0016181	.0401932	0	1
colony	422,715	.0129425	.1130268	0	1
comctry	422,715	.0005512	.0234712	0	1

-> custrict = 1

Variable	Obs	Mean	Std. Dev.	Min	Max
ltrade	4,077	10.57258	3.099791	-6.278147	17.74001
ldist	4,077	7.07738	1.011146	3.782556	9.350468
lrgdp	3,169	44.71219	3.059198	35.3876	52.50938
lrgdppc	3,169	14.51523	1.561307	11.80947	20.59256
comlang	4,077	.8474368	.35961	0	1
border	4,077	.1572234	.364056	0	1
regional	4,077	.073093	.260321	0	1
landl	4,077	.3056169	.5351257	0	2
island	4,077	.4422369	.7077391	0	2
lareap	4,077	23.17425	4.299664	10.43412	28.60295
comcol	4,077	.6610253	.4734193	0	1
curcol	4,077	.162129	.3686143	0	1
colony	4,077	.2251656	.4177425	0	1
comctry	4,077	.0885455	.2841214	0	1

* 2.2 restricting data to year 1997 from now on
* 2.A. obtain summary statistics for variable set A
summarize $A if year == 1997

```
. summarize $A if year == 1997
```

Variable	Obs	Mean	Std. Dev.	Min	Max
ltrade	14,105	10.14634	4.095246	-7.163213	23.86674
ldist	10,889	8.149808	.8190299	3.684131	9.421514
lrgdp	7,996	48.23964	2.833131	38.49837	58.01698
lrgdppc	7,996	16.22658	1.592804	11.23118	19.90353
comlang	14,105	.1386033	.3455444	0	1
border	14,105	.0180078	.1329841	0	1
regional	14,105	.0139667	.1173567	0	1
landl	10,890	.3421488	.5355533	0	2
island	10,890	.3606979	.5516053	0	2
lareap	10,890	23.26414	3.697372	5.521061	32.76884
comcol	14,105	.0702588	.2555917	0	1
curcol	14,105	.0008508	.0291565	0	1
colony	14,105	.0103509	.1012152	0	1
comctry	14,105	.0007799	.0279162	0	1

* 2.B restricting data to non missing of lrgdp set A

summarize $A if lrgdp !=. & year ==1997

```
. * 2.B restricting data to non missing of lrgdp set A
. summarize $A if lrgdp !=. & year ==1997
```

Variable	Obs	Mean	Std. Dev.	Min	Max
ltrade	7,996	9.771477	3.651089	-7.163213	20.70488
ldist	7,996	8.159072	.8054172	3.782556	9.421514
lrgdp	7,996	48.23964	2.833131	38.49837	58.01698
lrgdppc	7,996	16.22658	1.592804	11.23118	19.90353
comlang	7,996	.1885943	.3912105	0	1
border	7,996	.0251376	.1565527	0	1
regional	7,996	.0231366	.1503466	0	1
landl	7,996	.3528014	.5393517	0	2
island	7,996	.3289145	.5294285	0	2
lareap	7,996	23.82492	3.285805	11.39347	32.19601
comcol	7,996	.0975488	.2967221	0	1
curcol	7,996	0	0	0	0
colony	7,996	.0147574	.1205878	0	1
comctry	7,996	0	0	0	0

* 2.C obtain summary statistics for non missing lrgdp set A by currency union

by custrict, sort : summarize $A if lrgdp !=. & year ==1997

```
by custrict, sort : summarize $A if lrgdp !=. & year ==1997
```

-> custrict = 0

Variable	Obs	Mean	Std. Dev.	Min	Max
ltrade	7,913	9.783523	3.654291	-7.163213	20.70488
ldist	7,913	8.175889	.7876403	4.016798	9.421514
lrgdp	7,913	48.28449	2.802683	39.00064	58.01698
lrgdppc	7,913	16.24903	1.578818	11.23118	19.90353
comlang	7,913	.1810944	.3851207	0	1
border	7,913	.0230001	.1499132	0	1
regional	7,913	.0218628	.1462446	0	1
landl	7,913	.3506887	.5377366	0	2
island	7,913	.3290787	.5269625	0	2
lareap	7,913	23.83256	3.266342	11.45953	32.19601
comcol	7,913	.0898521	.2859879	0	1
curcol	7,913	0	0	0	0
colony	7,913	.0147858	.1207022	0	1
comctry	7,913	0	0	0	0

-> custrict = 1

Variable	Obs	Mean	Std. Dev.	Min	Max
ltrade	83	8.623031	3.144354	-5.870682	15.63367
ldist	83	6.555757	.8716879	3.782556	8.102106
lrgdp	83	43.96355	2.445333	38.49837	52.31883
lrgdppc	83	14.08631	1.470661	12.23975	19.15664
comlang	83	.9036145	.2969133	0	1
border	83	.2289157	.4226889	0	1
regional	83	.1445783	.353813	0	1
landl	83	.5542169	.648695	0	2
island	83	.313253	.7313154	0	2
lareap	83	23.09632	4.763322	11.39347	28.08278
comcol	83	.8313253	.3767407	0	1
curcol	83	0	0	0	0
colony	83	.0120482	.1097643	0	1
comctry	83	0	0	0	0

* 2.D what is std for current colony? what assumption it violates?

The std =zero. if we want to include this in reg analysis it will violate the third assumption of no perfect collinearity

* 2.E Run OLS regression of "Log real trade" on the explanatory variables identified in Table 2

* Run OLS egression of Log real Trade on B

reg ltrade $B if year == 1997

```
. reg ltrade custrict ldist lrgdp lrgdppc comlang border regional landl island lareap
> comcol colony if year == 1997
```

Source	SS	df	MS	Number of obs	=	7,996
				F(12, 7983)	=	1420.75
Model	72588.3656	12	6049.03047	Prob > F	=	0.0000
Residual	33988.5662	7,983	4.25761821	R-squared	=	0.6811
				Adj R-squared	=	0.6806
Total	106576.932	7,995	13.330448	Root MSE	=	2.0634

ltrade	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
custrict	.903533	.2422384	3.73	0.000	.4286825	1.378383
ldist	-1.273237	.0347425	-36.65	0.000	-1.341342	-1.205133
lrgdp	.9414817	.0145966	64.50	0.000	.9128686	.9700948
lrgdppc	.424044	.0222102	19.09	0.000	.3805062	.4675817
comlang	.4206994	.0659431	6.38	0.000	.2914337	.549965
border	.7448734	.1631055	4.57	0.000	.425144	1.064603
regional	.8932714	.1653786	5.40	0.000	.569086	1.217457
landl	-.3016932	.0461099	-6.54	0.000	-.3920806	-.2113058
island	-.0832614	.0550904	-1.51	0.131	-.1912531	.0247302
lareap	-.0931269	.0119216	-7.81	0.000	-.1164964	-.0697575
comcol	.3863664	.0903623	4.28	0.000	.2092327	.5635001
colony	1.310422	.1967422	6.66	0.000	.924756	1.696088
_cons	-29.97018	.5021204	-59.69	0.000	-30.95447	-28.98589

```
             (1)
VARIABLES        ltrade

custrict     0.904***
             (0.242)
ldist       -1.273***
             (0.0347)
lrgdp        0.941***
             (0.0146)
lrgdppc      0.424***
             (0.0222)
comlang      0.421***
             (0.0659)
border       0.745***
             (0.163)
regional     0.893***
             (0.165)
landl       -0.302***
             (0.0461)
island      -0.0833
             (0.0551)
lareap      -0.0931***
             (0.0119)
comcol       0.386***
             (0.0904)
colony       1.310***
             (0.197)
Constant    -29.97***
             (0.502)

Observations     7,996
R-squared        0.681
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
```

* 2.F Would you say that the independent variables explain much of the variation in log real trade in part (2.2E) above?

* we can say 68.11% of variation in dependent variable which is "log real trade" is explained by independent variables

Please include pdf of your do .le and well-labelled output (log) .le. In addition, prepare a summary table for results from part (2.2E) using outreg2 command. Your summary table should be similar in format to Table 2 of Glick and Rose (2002), but use 3 or 4 decimal points for coe¢ cient estimates and their standard errors.

VARIABLES	(1) ltrade
custrict	0.904***
	(0.242)
ldist	-1.273***
	(0.0347)
lrgdp	0.941***
	(0.0146)
lrgdppc	0.424***
	(0.0222)
comlang	0.421***
	(0.0659)
border	0.745***
	(0.163)
regional	0.893***
	(0.165)
landl	-0.302***
	(0.0461)
island	-0.0833
	(0.0551)
lareap	-0.0931***
	(0.0119)
comcol	0.386***
	(0.0904)
colony	1.310***
	(0.197)
Constant	-29.97***
	(0.502)
Observations	7,996
R-squared	0.681

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1