Research Methodology T02
-Analyzing Data-

Feb 11-18, 2016

Chihiro Shimizu, Yanjiang Zhang

Professor, Institute of Real Estate Studies
National University of Singapore

NUS National University of Singapore | RES Institute of Real Estate Studies

Outlines

- 1. Introduction: Software
- 2. Stata Demonstration by “code” (Panel Data with Simple OLS)
- 3. SPSS Demonstration by “clicking” (Realis Data)
  - Download & Import
  - Clean & descriptive analysis
  - Regression (One method of housing index)
- Result Report
- Homework
- 4. Stata Demonstration by “code” (As Appendix)
- 5. Eviews by “Code” (As Appendix)

Materials Used for each part

- 2. Stata Demonstration by "code" (Panel Data with Simple OLS)
  - Use Data T02_panel02.csv
- 3. SPSS Demonstration by "clicking" (Realis Data)
  - Use Data "T02_SPSS_Data.csv"
- 4. Stata Demonstration by "code" (As Appendix)
  - Use Data "panelEx.dta", where .dta is the format of data sheet
  - Use "Panel Routine.docx" as a guidance for normal panel data analysis
- 5. Eviews by "Code" (As Appendix)
  - Use Data "paneldata_10countries_1970_2013.csv"

1. Software or Program

- Advantage and disadvantage in Software or program.

- MS-Excel

- SPSS
- Eviews
- Stata
- SAS
- R

- MatLab

Advantage and disadvantage in Software or program

- Statistical Software
- Advantage and disadvantage in Software or program

SPSS	EViews	Stata	SAS	R
Social Science	Econ & Finance	Economics	Finance & Bio Stat	Bio Stat
Simple	Time Series	Regression & Clean Data	Big Data	Programing (Loop)
Basic	Basic	Professional	Professional	Professional

2. Stata

Pooled OLS regression:

- Load and set up the panel data (data cleaning omitted)
- Keep 5 countries
- Regression and report the results

2. Stata

- Load the data
  - Set up a folder for keeping materials

```
. * I set up a Folder named as "STATA" in disk E
. * cut and paste the data “T02_panel02.csv” into the folder
. cd "E:¥Stata"
```

  - Load the data

```
. import delimited "E:¥Stata¥T02_panel02.csv", clear
. edit
```

	code	t	hpi	gdppw	tpop	rolddep
1	JPN	1970	102.0395	38.67959	81.47498	.1021453
2	JPN	1971	111.5878	40.05632	82.53065	.1046405
3	JPN	1972	124.9006	42.95418	83.69485	.1072729
4	JPN	1973	148.6863	46.19217	84.39056	.1100685
5	JPN	1974	137.095	44.92836	86.01702	.1130661
6	JPN	1975	117.7916	45.72738	87.40533	.1162812
7	JPN	1976	110.2579	47.31423	88.05419	.1197443
8	JPN	1977	106.3665	49.00938	88.90685	.123417

45	AUS	1970	31.81757	54.15565	56.76794	.1311775
46	AUS	1971	33.50736	54.28769	58.71967	.1313649
47	AUS	1972	35.34076	55.21532	59.809	.1322995
48	AUS	1973	38.91178	55.61463	60.7304	.1337988
49	AUS	1974	41.31472	56.25379	62.28724	.1356029

2. Stata

- Keep the five selected countries
  - Keep The selected countries

. keep if code == "JPN" | code == "AUS" | code == "DNK" | code == "FIN" | code == "IRL"

  - Take a look at the countries

. tab code

```
. tab code

       code |      Freq.     Percent        Cum.
------------+-----------------------------------
        AUS |         44       20.00       20.00
        DNK |         44       20.00       40.00
        FIN |         44       20.00       60.00
        IRL |         44       20.00       80.00
        JPN |         44       20.00      100.00
------------+-----------------------------------
      Total |        220      100.00
```

2. Stata

- Descriptive statistics of variables
  - Summary of all

. sum hpi, detail

```
                              HPI
-------------------------------------------------------------
      Percentiles      Smallest
 1%     33.78582       31.81757
 5%      37.7318       33.50736
10%     39.45675       33.78582       Obs                 220
25%     49.02537       34.29065       Sum of Wgt.         220

50%     68.49821                      Mean           78.88929
                        Largest       Std. Dev.      36.06984
75%          100       167.6261
90%     135.2685       176.4367       Variance       1301.034
95%     150.0098        184.843       Skewness       .8583571
99%     176.4367       186.6228       Kurtosis        2.95724
```

  - Summary by time (country, et al)

. sort t

. by t: sum hpi, detail

```
-> t = 1971

                              HPI
-------------------------------------------------------------
      Percentiles      Smallest
 1%     33.50736       33.50736
 5%     33.50736       34.86092
10%     33.50736       53.51634       Obs                   5
25%     34.86092       58.68754       Sum of Wgt.           5

50%     53.51634                      Mean           58.43198
                        Largest       Std. Dev.      31.72773
75%     58.68754       34.86092
90%     111.5878       53.51634       Variance       1006.649
95%     111.5878       58.68754       Skewness        1.06325
99%     111.5878       111.5878       Kurtosis       2.676433
```

2. Stata

- Run the OLS
  - Regression

. reg hpi gdppw rolddep tpop

```
. reg hpi gdppw rolddep tpop

      Source |       SS           df       MS      Number of obs   =       220
-------------+----------------------------------   F(3, 216)       =     65.88
       Model |  136138.392         3  45379.464   Prob > F        =    0.0000
    Residual |  148787.975       216  688.833218   R-squared       =    0.4778
-------------+----------------------------------   Adj R-squared   =    0.4705
       Total |  284926.367       219  1301.03364   Root MSE        =    26.246

------------------------------------------------------------------------------
         hpi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       gdppw |   .3450735   .1213668     2.84   0.005     .1058586    .5842884
     rolddep |  -270.5863    45.4644    -5.95   0.000    -360.1969   -180.9756
        tpop |    2.31697   .2373012     9.76   0.000     1.849247    2.784692
       _cons |  -98.27931   15.58277    -6.31   0.000    -128.9931   -67.56557
------------------------------------------------------------------------------
```

	Coefficient	Std. Error	
Intercept	-98.279	15.583	***
GDPPW	0.345	0.121	***
OLDDEP	-270.586	45.464	***
TPOP	2.317	0.237	***
Adjusted R-squared:	0.471		
Number of observations:	220		

Note: Estimation method is OLS. Signs "***" indicate the estimated parameters are significant at 1%.

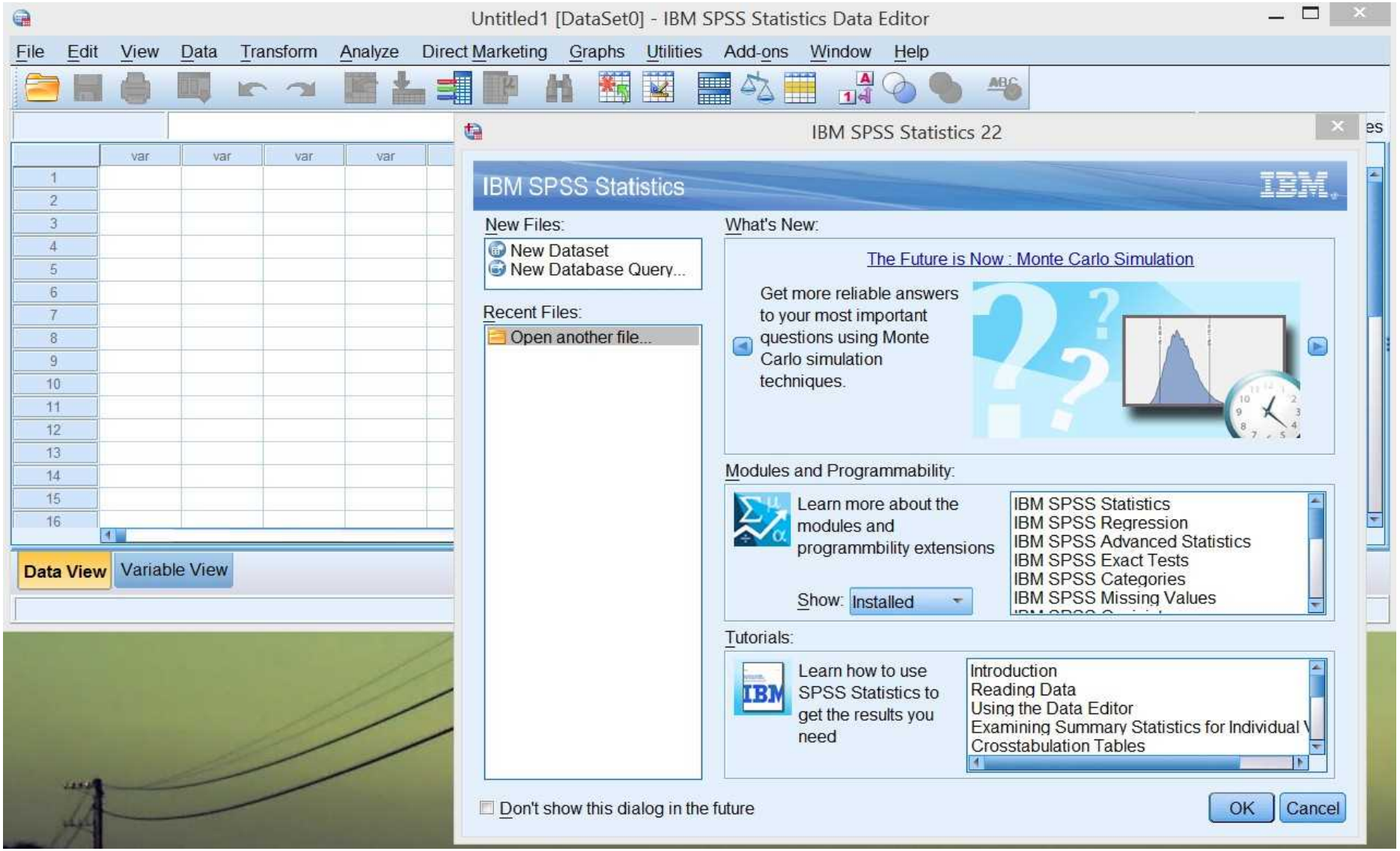
2. Stata

- What can stata do?
  - Regression
  - Data cleaning for small-median sample (several million observations)
  - The code (should demonstrate)

- How to learn? (should demonstrate)
  - 1. . help
  - 2. my PPT and materials
  - 3. http://data.princeton.edu/stata/ (follow the tutorials step by step)
  - 4. http://www.statalist.org/ (search whatever you want)
  - 5. use codes and commands from papers, for example, http://www.stata-journal.com/archives/
  - 6. google

Step 1: Finish 3, you will get a rough picture of stata;
Step 2: Finish 2, you can handle panel data regressions (an initial stage);
Step 3: learning by doing with the help of 1, 4, 5 and 6.

3. SPSS


Untitled1 [DataSet0] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
var var var var
Data View
Variable View
IBM SPSS Statistics 22
IBM SPSS Statistics
IBM
New Files:
New Dataset
New Database Query...
Recent Files:
Open another file...
What's New:
The Future is Now : Monte Carlo Simulation
Get more reliable answers to your most important questions using Monte Carlo simulation techniques.
Modules and Programmability:
Learn more about the modules and programmbility extensions
Show: Installed
IBM SPSS Statistics
IBM SPSS Regression
IBM SPSS Advanced Statistics
IBM SPSS Exact Tests
IBM SPSS Categories
IBM SPSS Missing Values
Tutorials:
Learn how to use SPSS Statistics to get the results you need
Introduction
Reading Data
Using the Data Editor
Examining Summary Statistics for Individual
Crosstabulation Tables
Don't show this dialog in the future
OK
Cancel

Import Data

- Data Source
  - Download from public source
  - Download from NUS database
  - Your own survey
  - Format: csv, excel, txt…

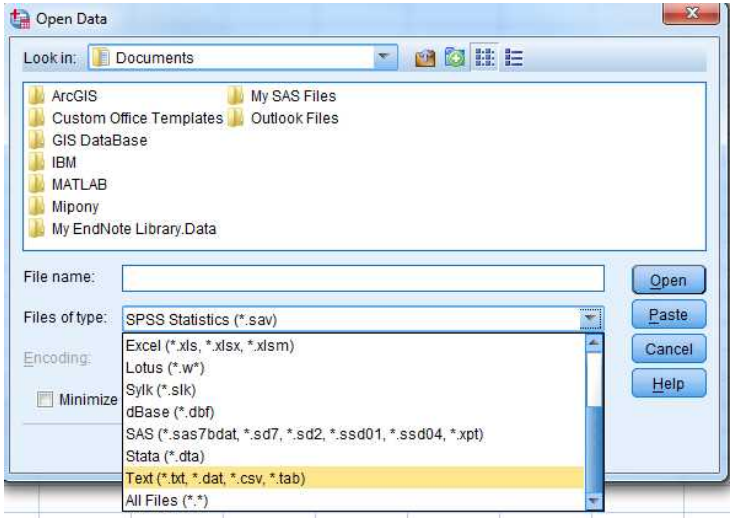
- Example for today – REALIS
  - Private housing transaction data

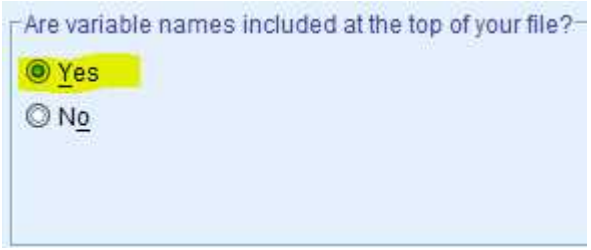
Import Data

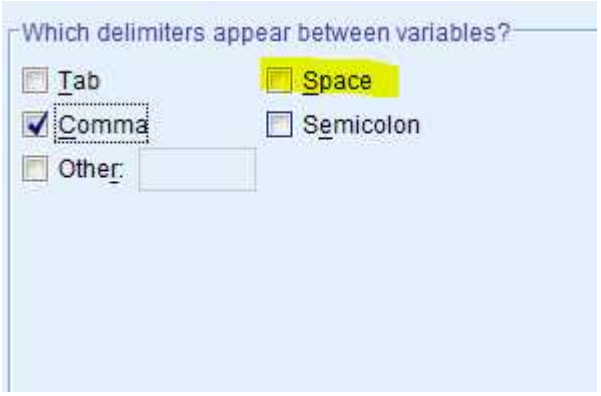
- REALIS Download
  - NUS Library – Database -- REALIS
  - Residential -- Transaction
  - Search -- Download/Print Record -- Download
  - Choose: 2015 Jan – 2015 Dec (Excluding Enblock)

Import Data

- Import Data







Clean Data

- Clean Data
  - Essential work before regression
  - Covert “text information” to computer readable
- Dummy Variable
- Time Variable
- Logarithm(log, ln) transformation

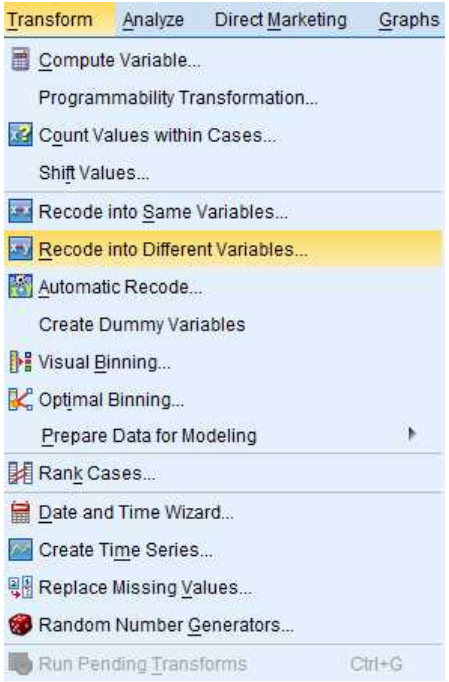
Clean Data

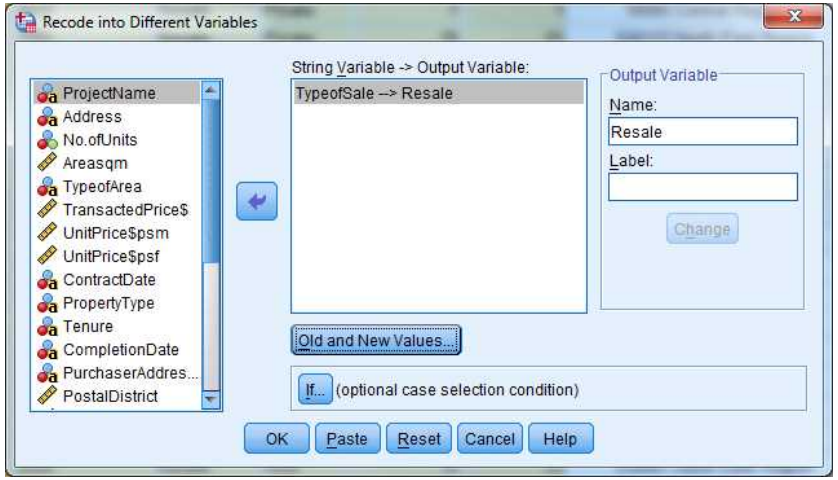
- Dummy Variable
  - Variable Type: numeric (Unit Price, Area)

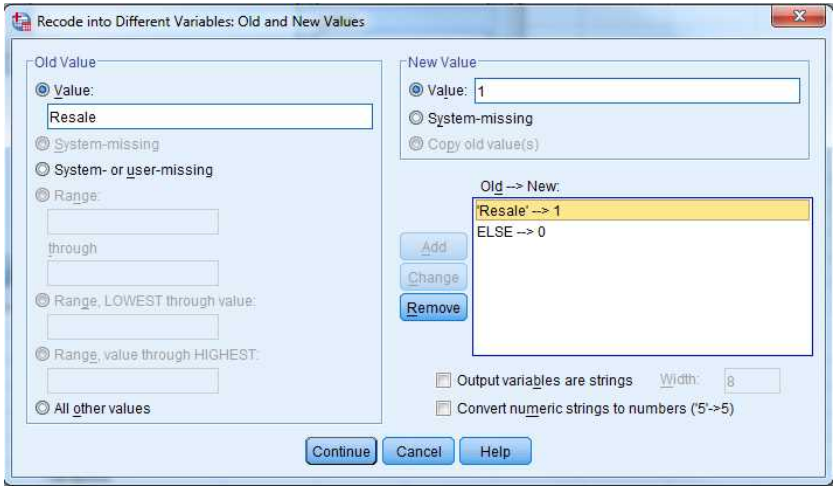
    string(Region, Type of Sales)
  - Variable in Regression should be numerical
    - Change string to dummy variable

Clean Data

- Two Value Variable:
  - Recode into Different Variables
  - “Resale”=1
  - “New sale”=0

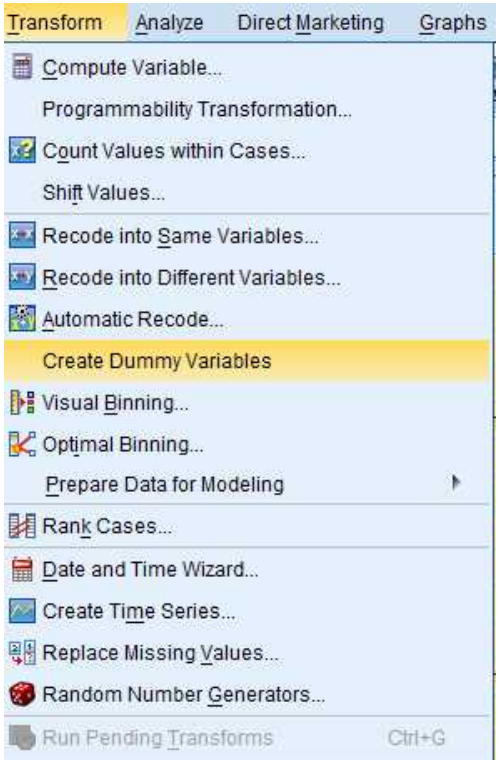


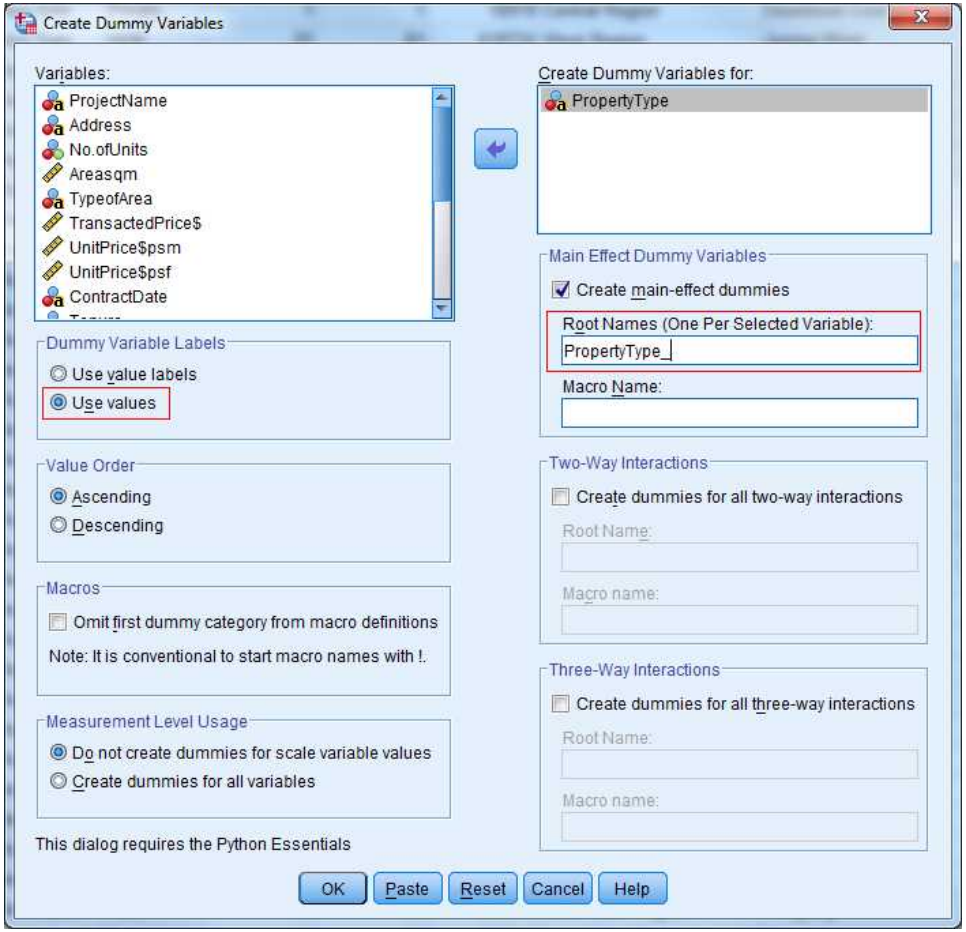




Clean Data

- Multiple Value Variable:
  - Create Dummy Variables





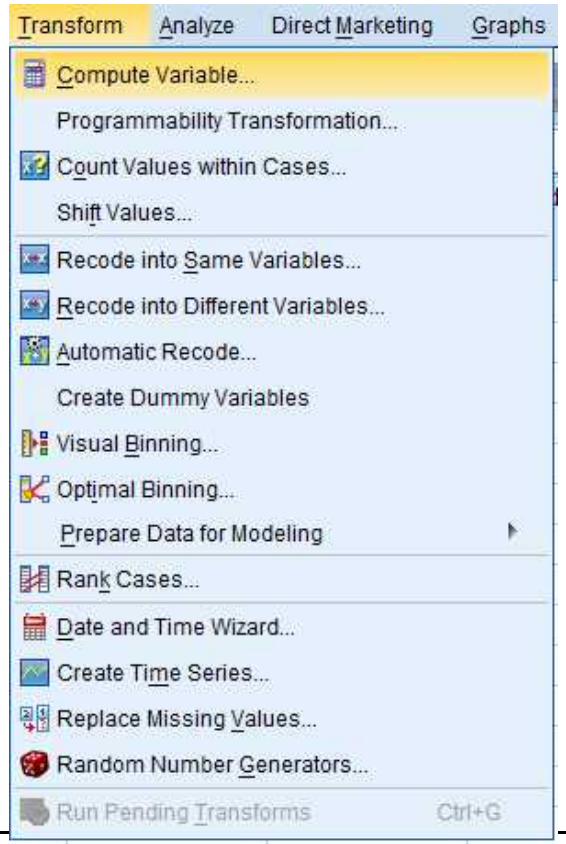
  - **Take a look at the "Variable View" to see the property type dummies**

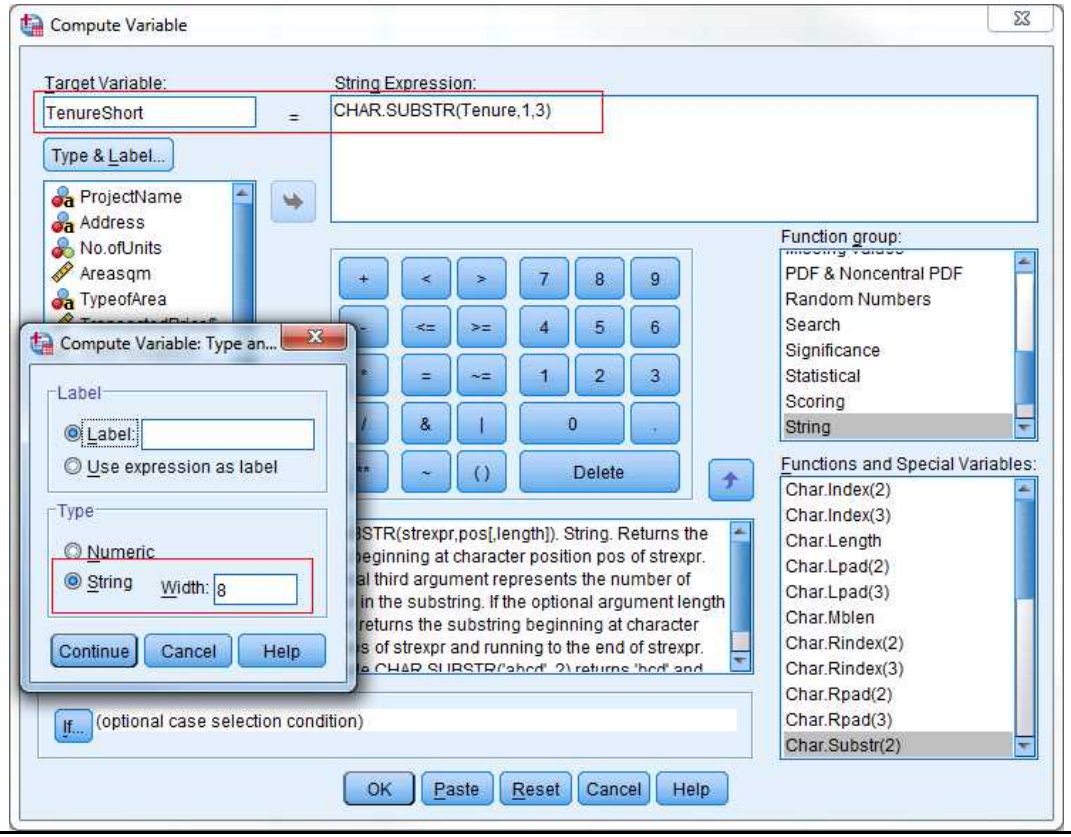
Clean Data

- Exercise
  1. Purchaser Address Indicator to dummy variable (HDB private)
  2. Planning Region to dummy variable (Central, East, West, North, NE)

Clean Data

- Complex Work:
  - How to change Tenure to Dummy Variable? (take a look at Tenure values first)





Clean Data

- Exercise

1. Completion Date to dummy variable (uncompleted, , unknown, before 2000, after 2000)

Tips: take the first 3 digits of the string, recode one by one, change string to be numeric (through variable view)

1. Transacted Price to logarithmic (ln(Transacted Price))

Tips: (through compute variable, arithmetic, Ln)

Clean Data

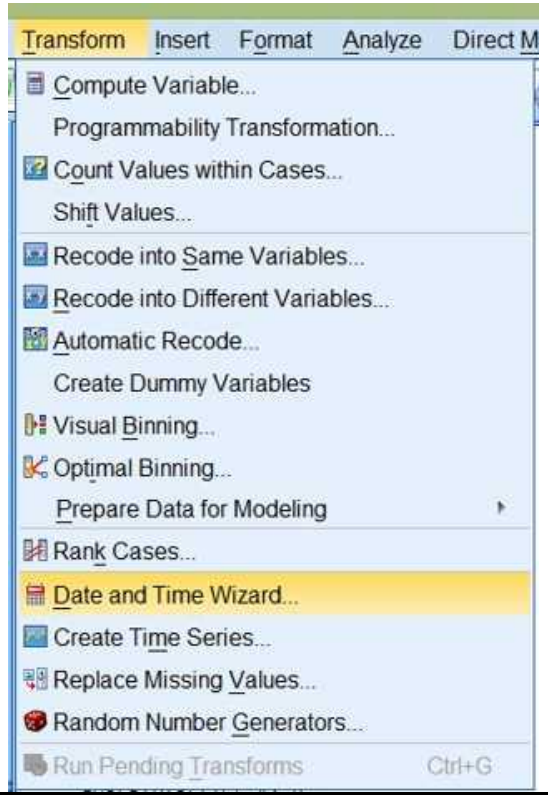
- Time Variable
  - Variable Type:

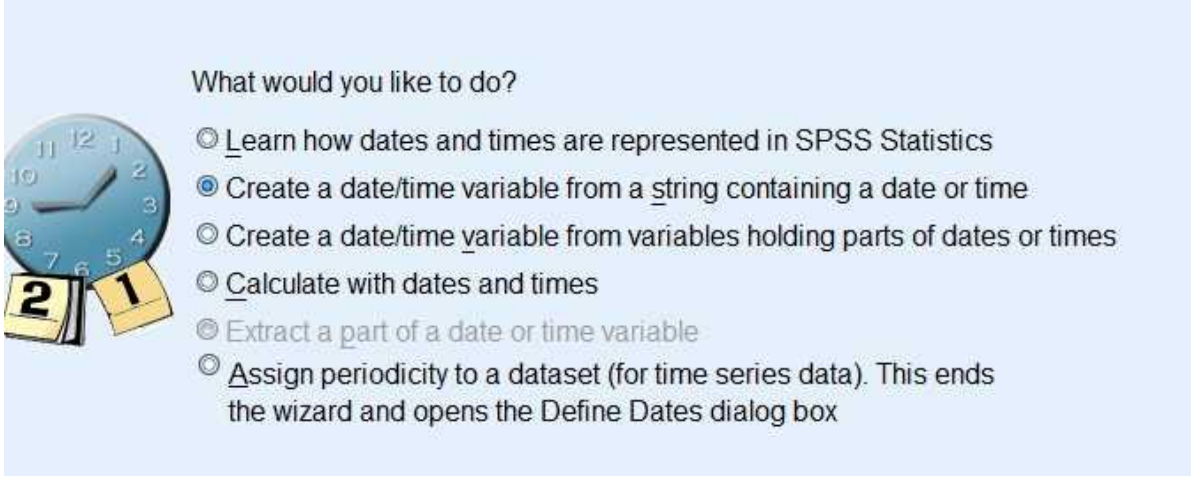
    string(SaleDate)
  - Housing prices have cyclical pattern
    - Change date format to Month, Quarter or even Year
  - Variable in Regression should be numerical
    - Change string to dummy variable

      (Coefficients of time dummies make up the price indexes)

Clean Data

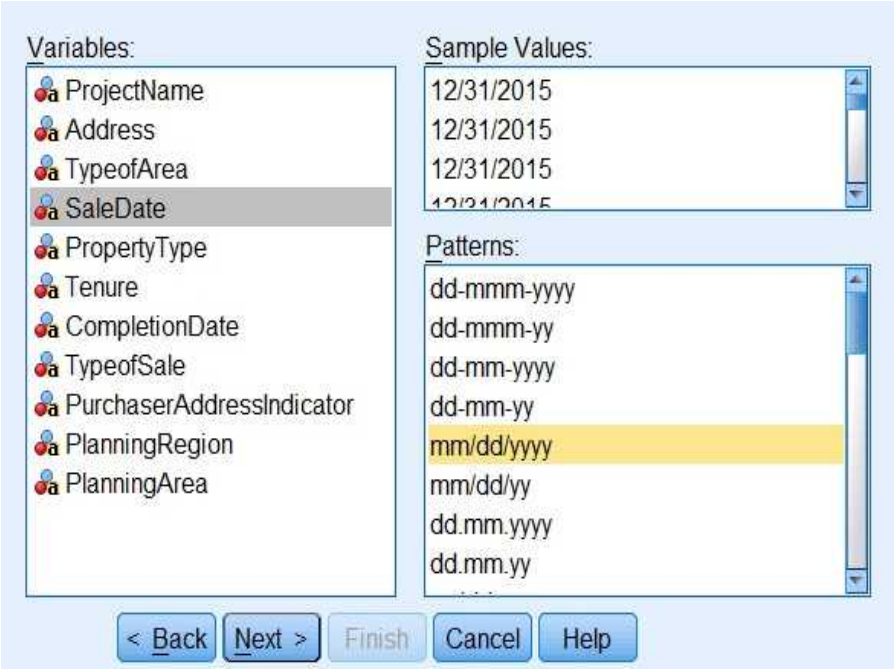
- Time Variable:
  - Date and Time Wizard
  - Create a new variable

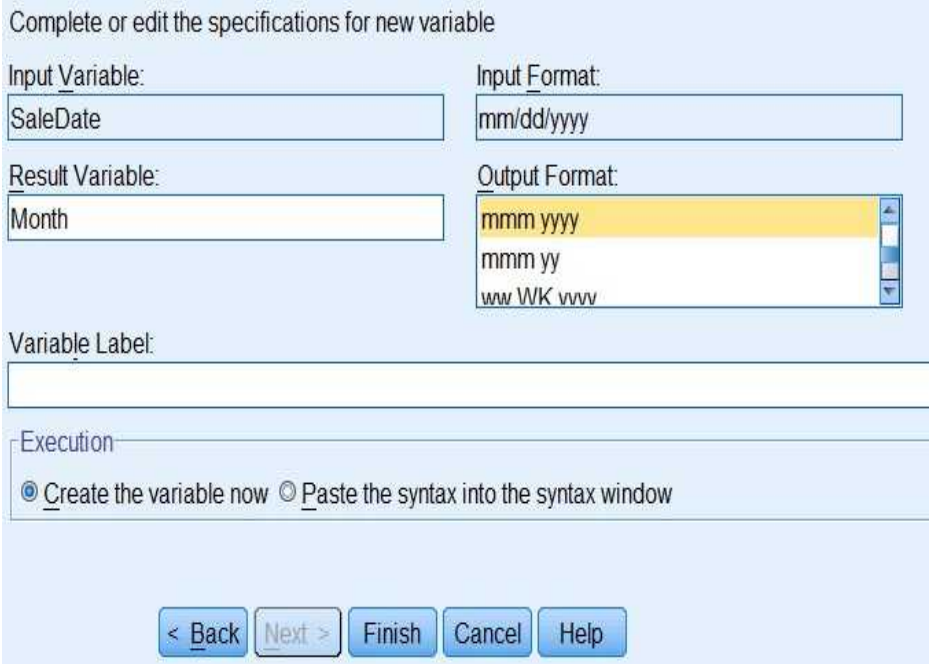




Clean Data

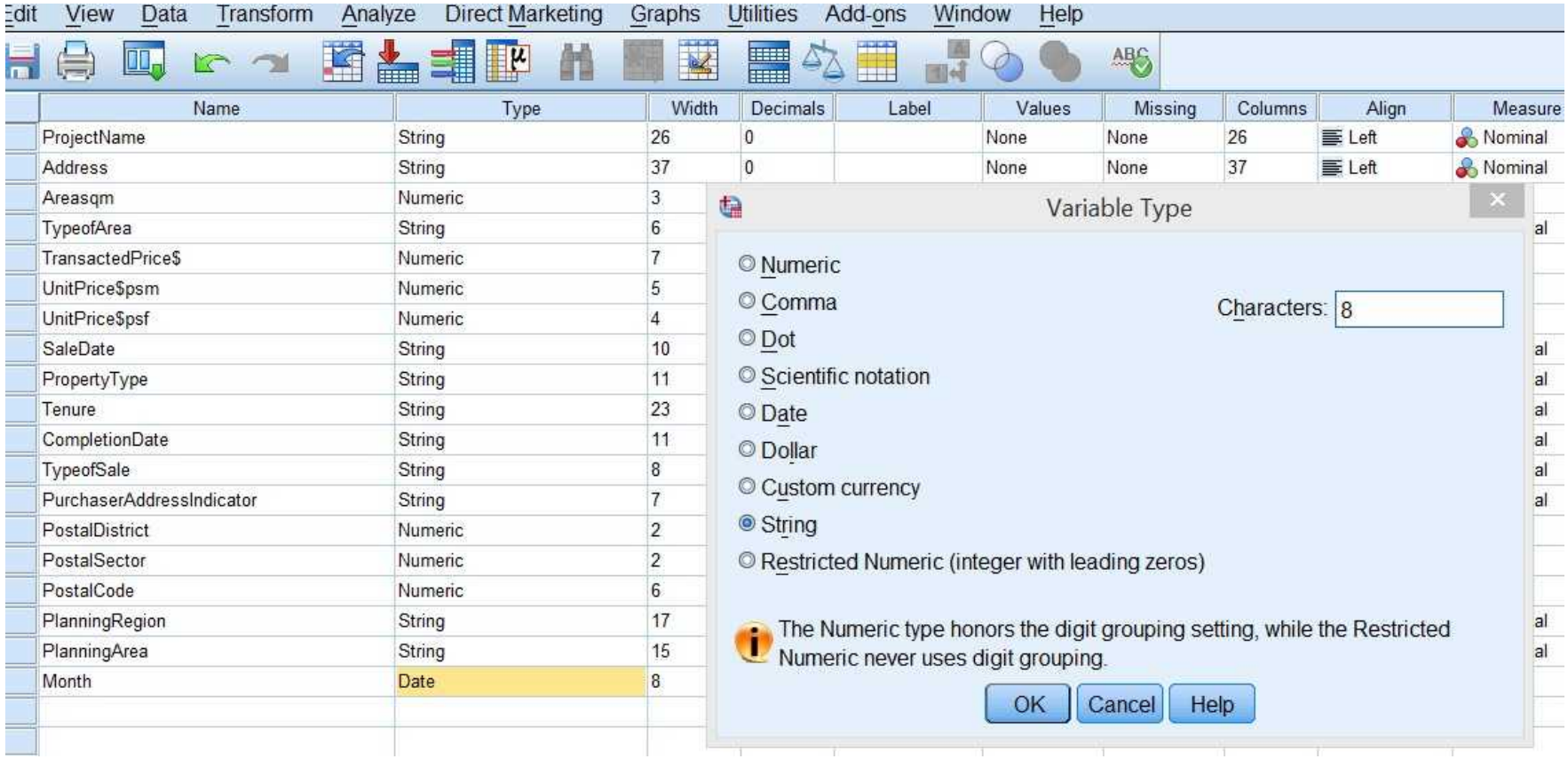
- Time Variable:
  - Date and Time Wizard
  - Create a new variable
  - **Define the format of input variable SaleDate "mm/dd/yyyy"**
  - **Define the format of output variable Month "mmm yyy"**





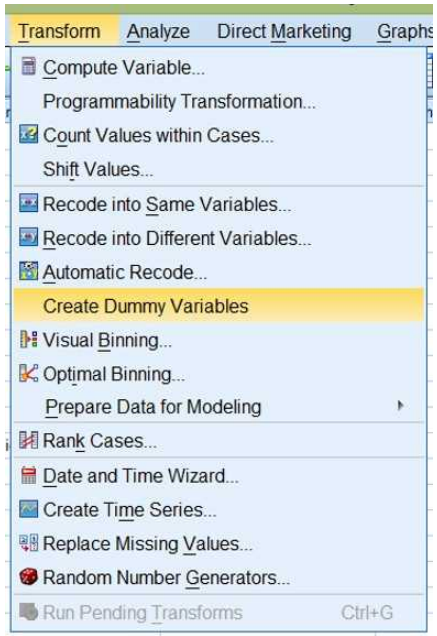
Clean Data

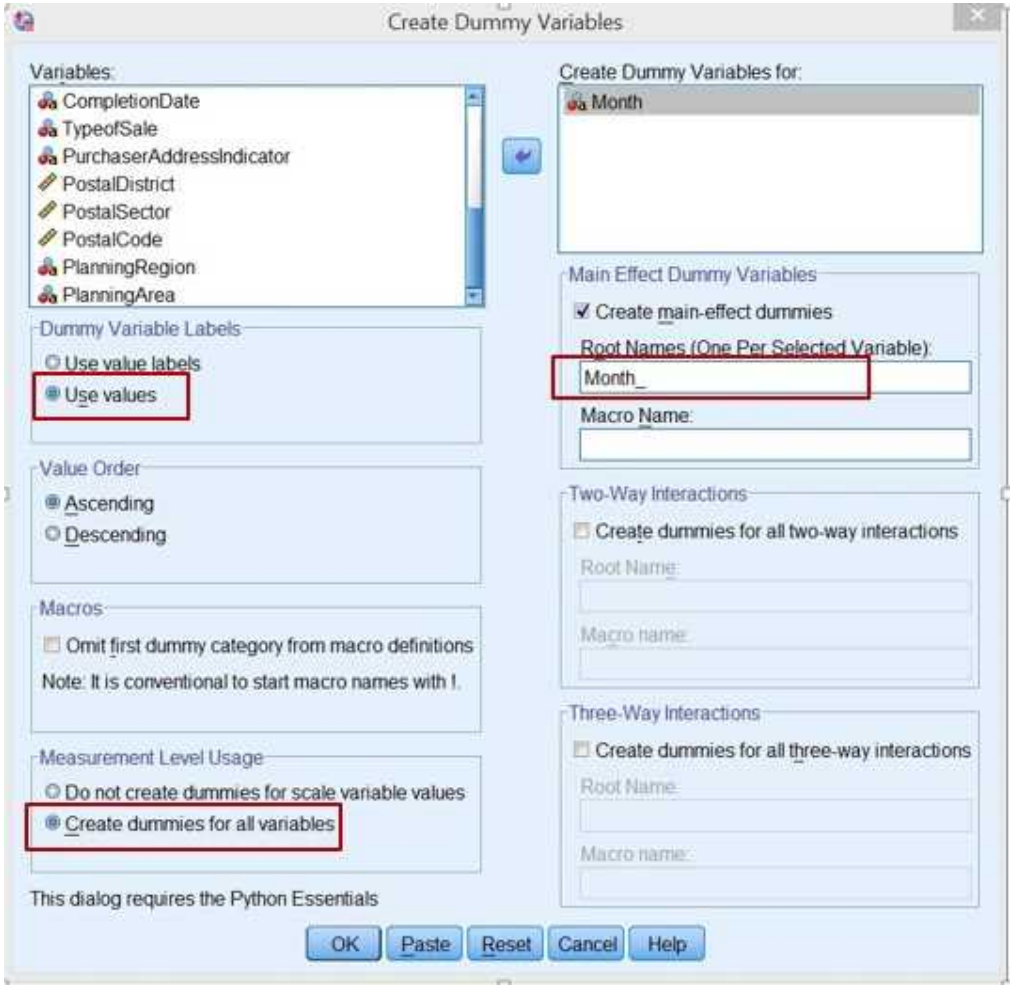
- Time Variable: create time dummy
  - Redefine the format of variable "Month"



Clean Data

- Time Variable: Create time dummies
  - Create Dummy Variables
  - Month ---Month_(Root Names)





Clean Data

- Time Variable: Create time dummies
  - The Dummy Variables created

Name	Type	Width	Decimals	Label
Month__1	Numeric	8	2	Month=APR 2015
Month__2	Numeric	8	2	Month=AUG 2015
Month__3	Numeric	8	2	Month=DEC 2015
Month__4	Numeric	8	2	Month=FEB 2015
Month__5	Numeric	8	2	Month=JAN 2015
Month__6	Numeric	8	2	Month=JUL 2015
Month__7	Numeric	8	2	Month=JUN 2015
Month__8	Numeric	8	2	Month=MAR 2015
Month__9	Numeric	8	2	Month=MAY 2015
Month__10	Numeric	8	2	Month=NOV 2015
Month__11	Numeric			
Month__12	Numeric			

PlanningArea	Month	Month__1	Month__2	Month__3	Month_
Clementi	DEC 2015	.00	.00	1.00	
Sengkang	DEC 2015	.00	.00	1.00	
Bukit Panjang	DEC 2015	.00	.00	1.00	
Bukit Panjang	DEC 2015	.00	.00	1.00	
Bukit Batok	DEC 2015	.00	.00	1.00	
Ang Mo Kio	DEC 2015	.00	.00	1.00	
Bishan	DEC 2015	.00	.00	1.00	
Bishan	DEC 2015	.00	.00	1.00	
Hougang	DEC 2015	.00	.00	1.00	

Description

- Frequencies (numerical & string)
- Descriptive (numerical)
- Graphs
- Correlations

Description

- Frequencies (numerical & string)

Frequencies

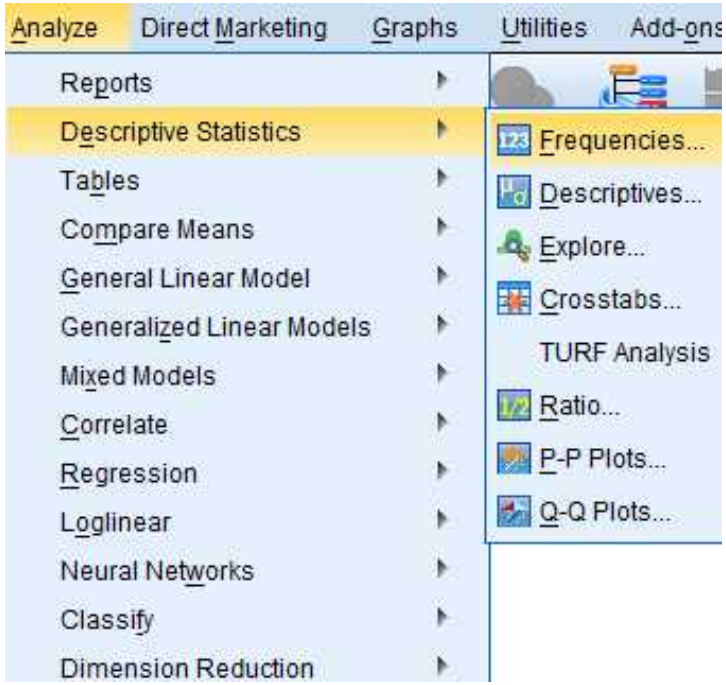
Statistics

PropertyType

N	Valid	13294
	Missing	0

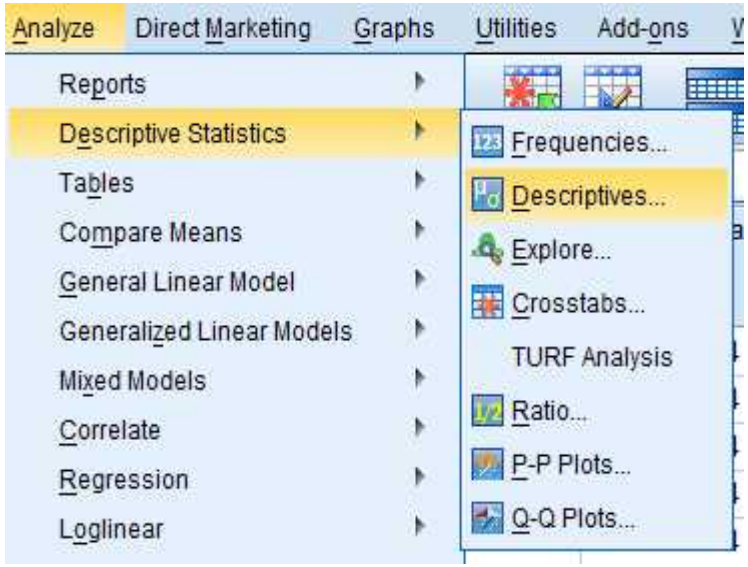
PropertyType

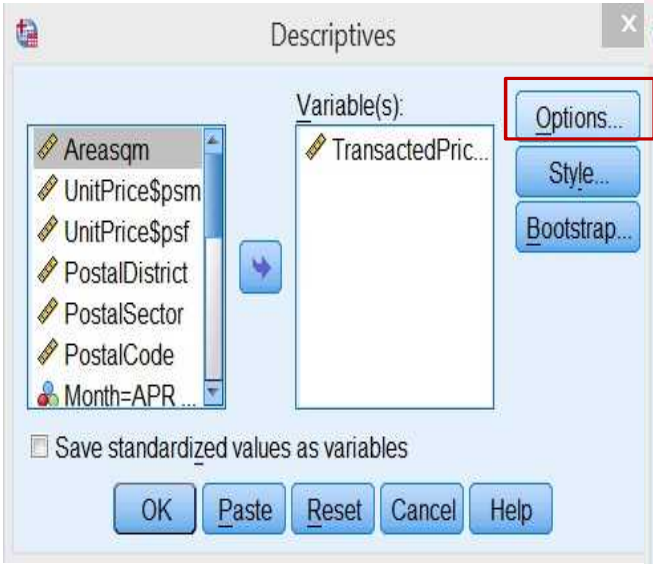
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Apartment	4951	37.2	37.2	37.2
	Condominium	7253	54.6	54.6	91.8
	Detached Ho	42	.3	.3	92.1
	Executive C	772	5.8	5.8	97.9
	Semi-Detach	95	.7	.7	98.6
	Terrace Hou	181	1.4	1.4	100.0
	Total	13294	100.0	100.0	

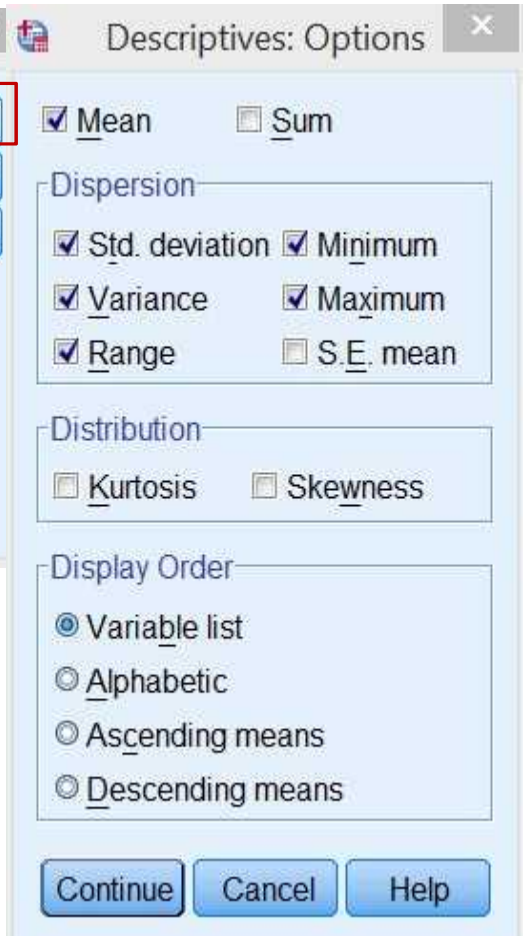


Description

- Descriptive (numerical)







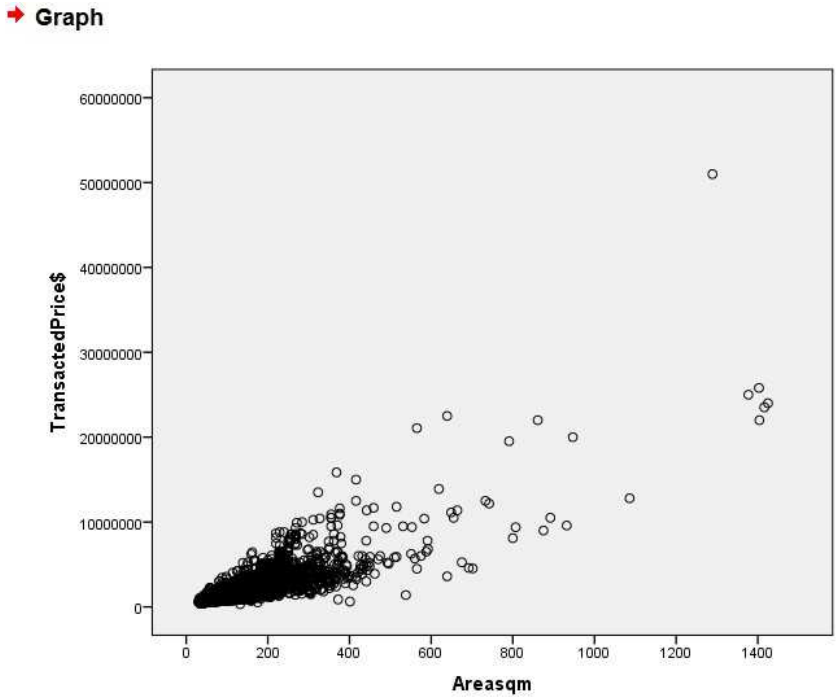
Descriptives

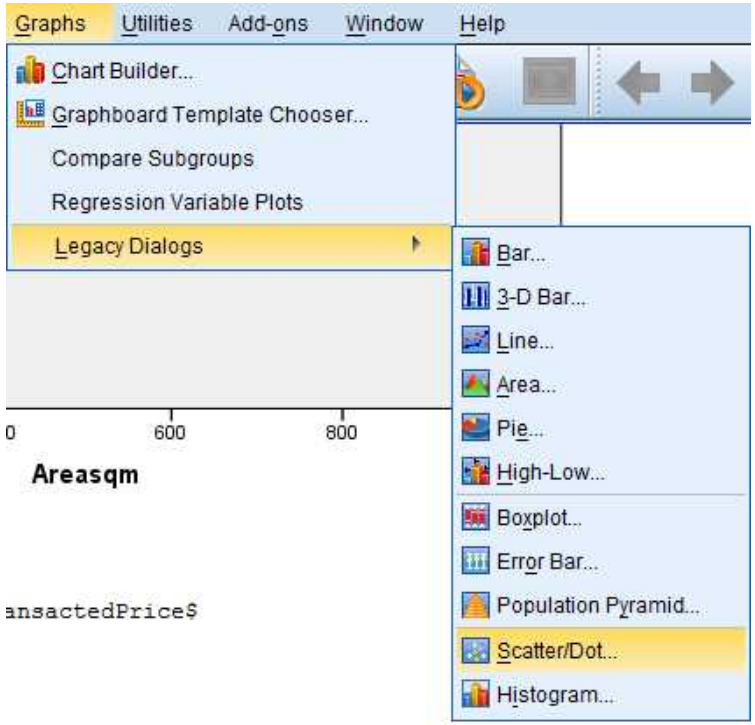
Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
TransactedPrice$	13294	50680000	320000	51000000	1395897.55	1262225.700	1.593E+12
Valid N (listwise)	13294						

Description

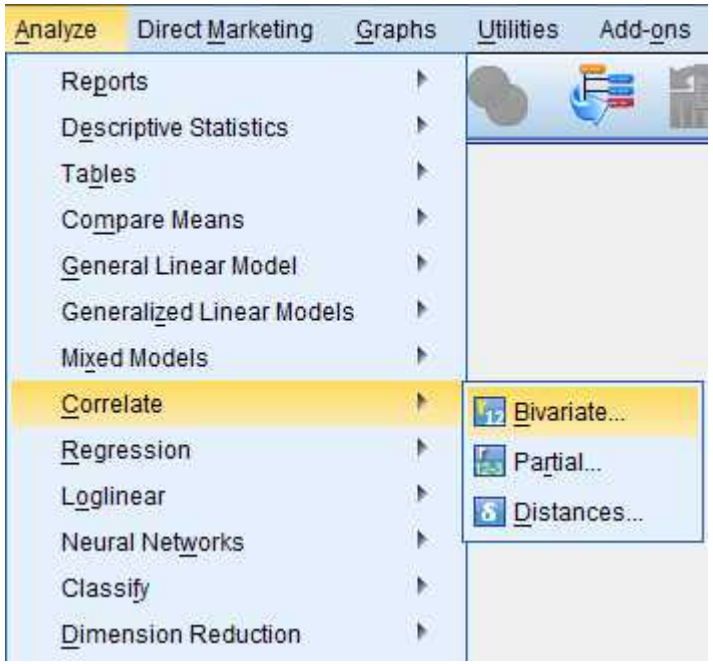
- Graphs





Description

- Correlations



Correlations

Correlations

		UnitPrice$psm	Areasqm
UnitPrice$psm	Pearson Correlation	1	-.038**
	Sig. (2-tailed)		.000
	N	13294	13294
Areasqm	Pearson Correlation	-.038**	1
	Sig. (2-tailed)	.000	
	N	13294	13294

**. Correlation is significant at the 0.01 level (2-tailed).

Description

- Exercise
  - Frequencies – Planning Region
  - Descriptive – Unit Price
  - Graphs – ln(Transactd Price) with Area
  - Correlation – Unitprice$psm with Type of Sale

Regression

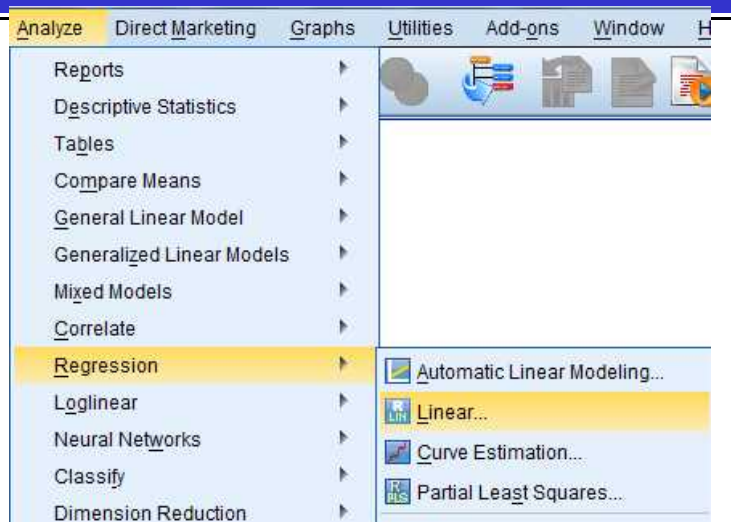
- Cross-section
  - Simple Linear Regression

$$Y = \beta_0 + \beta_1 x_1 + \varepsilon$$

  - Multiple Linear Regression

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \varepsilon$$

Regression



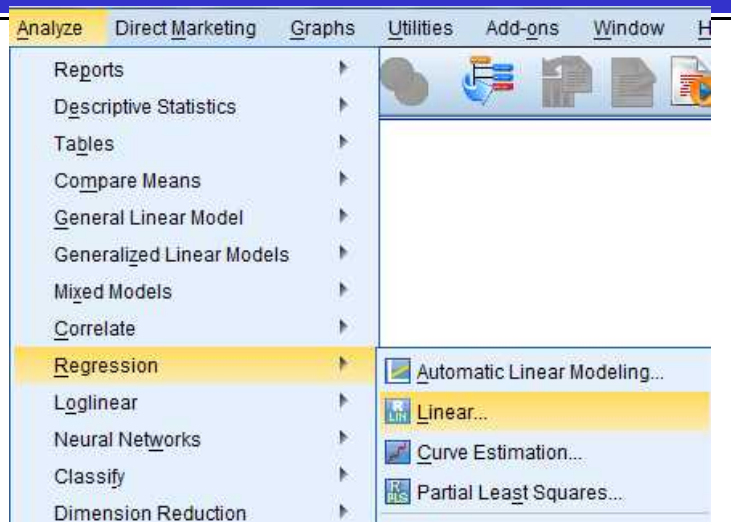
– Simple Linear Regression

$$Y = \beta_0 + \beta_1 x_1 + \varepsilon$$

$$\widehat{TransactedPrice} = b_0 + b_1 Areasqm$$

Regression

- Simple Linear Regression



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.821[a]	.674	.674	721186.700

a. Predictors: (Constant), Areasqm

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.427E+16	1	1.427E+16	27427.424	.000[b]
	Residual	6.913E+15	13292	5.201E+11		
	Total	2.118E+16	13293			

a. Dependent Variable: TransactedPrice$

b. Predictors: (Constant), Areasqm

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-203949.175	11508.390		-17.722	.000
	Areasqm	15266.734	92.184	.821	165.612	.000

a. Dependent Variable: TransactedPrice$

Regression

– Multiple Linear Regression

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \varepsilon$$

$$\widehat{LnPrice} = b_0 + b_1 Areasqm + b_2 Resale + b_3 Condo + b_4 Det + b_5 Excutive + b_6 Semi\ Deta + b_7 Terrace$$

$$+\alpha_2 Feb + \alpha_3 Mar + \alpha_4 Apr + \alpha_5 May + \alpha_6 June + \alpha_7 July + \alpha_8 Aug + \alpha_9 Sep + \alpha_{10} Oct + \alpha_{11} Nov + \alpha_{12} Dec$$

Regression

- Multiple Linear Regression

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	13.341	.012		1090.038	.000
	Areasqm	.006	.000	.859	124.745	.000
	Resale	.022	.006	.021	3.799	.000
	PropertyType=Condominium	.069	.006	.068	12.015	.000
	PropertyType=Detached Ho	-1.897	.054	-.211	-35.048	.000
	PropertyType=Executive C	-.263	.013	-.122	-20.734	.000
	PropertyType=Semi-Detach	-.331	.033	-.055	-10.087	.000
	PropertyType=Terrace Hou	.034	.023	.008	1.451	.147
	Month=FEB 2015	-.013	.014	-.007	-.935	.350
	Month=MAR 2015	-.045	.014	-.024	-3.304	.001
	Month=APR 2015	-.040	.013	-.024	-3.074	.002
	Month=MAY 2015	-.017	.014	-.010	-1.282	.200
	Month=JUN 2015	.014	.014	.008	.993	.321
	Month=JUL 2015	-.189	.013	-.136	-14.820	.000
	Month=AUG 2015	-.039	.014	-.020	-2.716	.007
	Month=SEP 2015	-.037	.015	-.017	-2.426	.015
	Month=OCT 2015	-.015	.014	-.008	-1.086	.277
	Month=NOV 2015	-.007	.014	-.004	-.470	.638
	Month=DEC 2015	-.016	.015	-.007	-1.054	.292

a. Dependent Variable: LnPrice

Dependent:
LnPrice

Independent(s):
- Areasqm
- Resale
- PropertyType=Condominium [PropertyType__2]
- PropertyType=Detached Ho [PropertyType__3]
- PropertyType=Executive C [PropertyType__4]
- PropertyType=Semi-Detach [PropertyType__5]
- PropertyType=Terrace Hou [PropertyType__6]
- Month=FEB 2015 [Month__4]
- Month=MAR 2015 [Month__8]
- Month=APR 2015 [Month__1]
- Month=MAY 2015 [Month__9]
- Month=JUN 2015 [Month__7]
- Month=JUL 2015 [Month__6]
- Month=AUG 2015 [Month__2]
- Month=SEP 2015 [Month__12]
- Month=OCT 2015 [Month__11]
- Month=NOV 2015 [Month__10]
- Month=DEC 2015 [Month__3]

Jan	=exp(0)*100
Feb	=exp(-0.013)*100
Mar	=exp(-0.045)*100
April	=exp(-0.040)*100
May	=exp(-0.017)*100
June	=exp(0.014)*100
July	=exp(-0.189)*100
Aug	=exp(-0.039)*100
Sep	=exp(-0.037)*100
Oct	=exp(-0.015)*100
Nov	=exp(-0.007)*100
Dec	=exp(-0.016)*100

Price Index

How to arrange for thesis?

- Report writing and Presentation.

4. Report writing and Presentation

Summary Statistics

	HPI	GDPPW	ROLDDEP	TPOP
Mean	77.77026	79.01426	0.19404	88.36226
Median	71.55661	80.21578	0.197377	92.22301
Maximum	186.6228	113.5407	0.40532	104.9814
Minimum	23.52028	38.67959	0.053268	43.48536
Std. Dev.	32.87626	18.25599	0.064853	12.41079
Skewness	0.682771	-0.13352	-0.25249	-1.38427
Kurtosis	2.869902	1.790482	3.415605	4.462461
Jarque-Bera	34.49655	28.12789	7.84182	179.7326
Probability	0	0.000001	0.019823	0
Sum	34218.91	34766.28	85.37779	38879.39
Sum Sq. Dev.	474492.3	146310.5	1.846391	67618.14
Observations	440	440	440	440

	HPI	GDPPW	OLDDEP	TPOP
Mean	77.77	79.01	0.19	88.36
Median	71.56	80.22	0.2	92.22
Maximum	186.62	113.54	0.41	104.98
Minimum	23.52	38.68	0.05	43.49
Std. Dev.	32.88	18.26	0.06	12.41
Observations	440	440	440	440

Regressions:

```
Random-effects GLS regression                   Number of obs     =      1,591
Group variable: i                               Number of groups  =         37

R-sq:                                           Obs per group:
     within  = 0.1032                                         min =         43
     between = 0.2409                                         avg =       43.0
     overall = 0.1412                                         max =         43

                                                Wald chi2(1)      =     189.72
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000

------------------------------------------------------------------------------
    D.ln_CPI |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rDEP |   .4537882   .0329451    13.77   0.000     .3892169    .5183594
       _cons |   -.175332   .0214788    -8.16   0.000    -.2174296   -.1332343
-------------+----------------------------------------------------------------
     sigma_u |  .06224765
     sigma_e |  .11020992
         rho |  .24185576   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

. outreg2 using 1.doc / or by hand

VARIABLES	(1) D.ln_CPI
rDEP	0.454***
	(0.0329)
Constant	-0.175***
	(0.0215)
Observations	1,591
Number of i	37

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

Regressions:

Dependent Variable: DLN_HPI

Method: Panel Least Squares

Date: 01/21/16 Time: 16:25

Sample (adjusted): 1971 2013

Periods included: 43

Cross-sections included: 10

Total panel (balanced) observations: 430

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.016308	0.006042	-2.698955	0.0072
DLN_GDPPW	1.474530	0.129994	11.34304	0.0000
DLN_OLDDEP	-0.101931	0.211673	-0.481549	0.6304
DLN_TPOP	1.258830	0.444428	2.832470	0.0048

R-squared	0.232792	Mean dependent var	0.015444
Adjusted R-squared	0.227389	S.D. dependent var	0.067392
S.E. of regression	0.059237	Akaike info criterion	-2.805294
Sum squared resid	1.494825	Schwarz criterion	-2.767491
Log likelihood	607.1382	Hannan-Quinn criter.	-2.790367
F-statistic	43.08659	Durbin-Watson stat	1.038223
Prob(F-statistic)	0.000000		

	Coefficient	Std. Error	
Intercept	-0.016	0.006	***
⊿lnGDPPW	1.475	0.13	***
⊿lnOLDDEP	-0.102	0.212	
⊿lnTPOP	1.259	0.444	***
Adjusted R-squared:	0.227		
Number of observations:	430		

Note: Estimation method is OLS. Signs "***" indicate the estimated parameters are significant at 1%.

Homework

- Q1. Find the annual "GDP per capita (constant)" and "Electric Power consumption" of 40 countries for 2000-2010;
- Q2. Use SPSS to show a correlation between "GDP per capita (constant)" and "Electric Power consumption" for the year of 2010;
- Q3. Use SPSS to show the relationship between "GDP per capita (constant)" and "Electric Power consumption" for the year of 2010 by OLS;
- Q4. Use STATA to show the relationship between "GDP per capita (constant)" and "Electric Power consumption" for the period between 2000-2010 (As panel data).

Upload your work to Workbin within 1 word doc, which should be named in the form "Your Group No._Your Student ID_Your Name_HW1". It should only include:

1. A list of country names for Q1;
2. Correlation results for Q2;
3. OLS results for Q3;
4. Regression results for Q4 (Simple OLS).

Deadline: within 7 days after this class.

The End & please take a look at the appendix

4. Appendix. Stata Routine for panel regression

Panel Regressions: Depend Ratio & CPI

- Load and set up the panel data (data cleaning omitted)
- Fixed effect model (FE) and the time effect
- Random effect model (RE) and the time effect
- Which one is more proper?

4. Appendix. Stata Routine for panel regression

- Load and set up the panel data (data cleaning omitted)

```
. cd "location of file folder where you work on"
. use panelEx.dta, clear
. xtset i t
. Order code i t ln_CPI rDEP
. edit
```

	code	i	t	ln_CPI	rDEP	complete_c~s	count	CPI
1	JPN	1	1970	3.4866103	.45251354	1	44	32.675
2	JPN	1	1971	3.5481796	.45531793	1	44	34.75
3	JPN	1	1972	3.5954841	.45979617	1	44	36.433333
4	JPN	1	1973	3.7054088	.46519511	1	44	40.666667
5	JPN	1	1974	3.9138547	.47059182	1	44	50.091667
6	JPN	1	1975	4.0252029	.47529296	1	44	55.991667
7	JPN	1	1976	4.1149638	.4790301	1	44	61.25
8	JPN	1	1977	4.1931839	.48179924	1	44	66.233333
9	JPN	1	1978	4.2344688	.48352008	1	44	69.025
10	JPN	1	1979	4.2707459	.48419762	1	44	71.575
11	JPN	1	1980	4.3459676	.48384463	1	44	77.166667
12	JPN	1	1981	4.3939346	.48251353	1	44	80.958333
13	JPN	1	1982	4.4207464	.48019939	1	44	83.158333
14	JPN	1	1983	4.4393124	.47683338	1	44	84.716667
15	JPN	1	1984	4.4619732	.47233398	1	44	86.658333
16	JPN	1	1985	4.4821547	.46674666	1	44	88.425

Variables

Filter variables here

Name	Label
i	i
t	t
ln_CPI	ln_CPI
rDEP	rDEP
complete_...	complete_cas..
count	count
CPI	CPI
GDP	GDP
GDPPW	GDPPW
TPOP	TPOP
YNGPOP	YNGPOP

4. Appendix. Stata Routine for panel regression

- Fixed effect model (FE) and the time effect
  - Fit FE model with no time effect

. xtreg D.ln_CPI rDEP, fe

. estimates store xt_fe0

```
Fixed-effects (within) regression               Number of obs      =      1,591
Group variable: i                               Number of groups   =         37

R-sq:                                           Obs per group:
     within  = 0.1032                                         min =         43
     between = 0.2409                                         avg =       43.0
     overall = 0.1412                                         max =         43

                                                F(1,1553)          =     178.79
corr(u_i, Xb)  = -0.0667                        Prob > F           =     0.0000

------------------------------------------------------------------------------
    D.ln_CPI |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rDEP |   .4586709   .0343024    13.37   0.000      .391387    .5259547
       _cons |  -.1781008   .0196473    -9.06   0.000     -.216639   -.1395627
-------------+----------------------------------------------------------------
     sigma_u |  .06383282
     sigma_e |  .11020992
         rho |  .25119693   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(36, 1553) = 14.36                   Prob > F = 0.0000
```

4. Appendix. Stata Routine for panel regression

- Fixed effect model (FE) and the time effect
  - Fit FE model with time effect

```
. tab t, gen(Year_)
. drop Year_1
```

Year_2	Year_3	Year_4	Year_5	Year_
0	0	0	0	
1	0	0	0	
0	1	0	0	
0	0	1	0	
0	0	0	1	
0	0	0	0	
0	0	0	0	
0	0	0	0	
0	0	0	0	
0	0	0	0	
0	0	0	0	
0	0	0	0	
0	0	0	0	
0	0	0	0	
0	0	0	0	
0	0	0	0	
0	0	0	0	
0	0	0	0	
0	0	0	0	
0	0	0	0	

Variables

Filter variables here

Name	Label
Year_2	t== 1971.00...
Year_3	t== 1972.00...
Year_4	t== 1973.00...
Year_5	t== 1974.00...
Year_6	t== 1975.00...
Year_7	t== 1976.00...
Year_8	t== 1977.00...
Year_9	t== 1978.00...
Year_10	t== 1979.00...
Year_11	t== 1980.00...
Year_12	t== 1981.00...
Year_13	t== 1982.00...
Year_14	t== 1983.00...
Year_15	t== 1984.00

4. Appendix. Stata Routine for panel regression

- Fixed effect model (FE) and the time effect
  - Fit FE model with time effect

. xtreg D.ln_CPI rDEP Year_*, fe

. estimates store xt_fe1

(. set more off)

```
Fixed-effects (within) regression               Number of obs      =      1,591
Group variable: i                               Number of groups   =         37

R-sq:                                           Obs per group:
     within  = 0.1974                                         min =         43
     between = 0.2409                                         avg =       43.0
     overall = 0.1928                                         max =         43

                                                F(43,1511)         =       8.64
corr(u_i, Xb)  = 0.0885                         Prob > F           =     0.0000

------------------------------------------------------------------------------
    D.ln_CPI |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       rDEP  |   .1879095   .0479001     3.92   0.000     .0939517    .2818673
     Year_2  |   .0109764    .026035     0.42   0.673    -.0400922     .062045
     Year_3  |   .0241612   .0259539     0.93   0.352    -.0267483    .0750706
     Year_4  |   .0955003   .0258667     3.69   0.000     .0447619    .1462386
             |
     Year_41 |   .0070254   .0245786     0.29   0.775    -.0411863    .0552371
     Year_42 |   .0154131   .0245777     0.63   0.531     -.032797    .0636231
     Year_43 |   .0062214   .0245766     0.25   0.800    -.0419864    .0544293
     Year_44 |          0  (omitted)
       _cons |  -.0738536   .0298744    -2.47   0.014    -.1324534   -.0152539
-------------+----------------------------------------------------------------
     sigma_u |  .06629367
     sigma_e |  .10570513
         rho |  .28229258   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(36, 1511) = 15.93                   Prob > F = 0.0000
```

4. Appendix. Stata Routine for panel regression

- Fixed effect model (FE) and the time effect
  - Fit FE model with time effect

. test Year_2 Year_3 Year_4 Year_5 Year_6 Year_7 Year_8 Year_9 Year_10 Year_11…Year_41 Year_42 Year_43 Year_44

```
(36)  Year_37 = 0
(37)  Year_38 = 0
(38)  Year_39 = 0
(39)  Year_40 = 0
(40)  Year_41 = 0
(41)  Year_42 = 0
(42)  Year_43 = 0
(43)  o.Year_44 = 0
      Constraint 43 dropped

      F( 42,  1511) =    4.22
           Prob > F =    0.0000
```

Time effect should be considered.

4. Appendix. Stata Routine for panel regression

- Random effect model (RE) and the time effect
  - Fit RE model with no time effect

. xtreg D.ln_CPI rDEP, re

. estimates store xt_re0

```
Random-effects GLS regression                   Number of obs     =      1,591
Group variable: i                               Number of groups  =         37

R-sq:                                           Obs per group:
     within  = 0.1032                                         min =         43
     between = 0.2409                                         avg =       43.0
     overall = 0.1412                                         max =         43

                                                Wald chi2(1)      =     189.72
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000

------------------------------------------------------------------------------
    D.ln_CPI |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       rDEP  |   .4537882   .0329451    13.77   0.000     .3892169    .5183594
       _cons |   -.175332   .0214788    -8.16   0.000    -.2174296   -.1332343
-------------+----------------------------------------------------------------
     sigma_u |  .06224765
     sigma_e |  .11020992
         rho |  .24185576   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

4. Appendix. Stata Routine for panel regression

- Random effect model (RE) and the time effect
  - Fit RE model with time effect

. xtreg D.ln_CPI rDEP Year_*, re

. estimates store xt_re1

```
Random-effects GLS regression                   Number of obs     =      1,591
Group variable: i                               Number of groups  =         37

R-sq:                                           Obs per group:
     within  = 0.1972                                         min =         43
     between = 0.2409                                         avg =       43.0
     overall = 0.1972                                         max =         43

                                                Wald chi2(43)     =     379.62
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000

------------------------------------------------------------------------------
    D.ln_CPI |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       rDEP  |   .2169855   .0444403     4.88   0.000     .1298842    .3040869
     Year_2  |   .0057601   .0258491     0.22   0.824    -.0449032    .0564234
     Year_3  |   .0190961   .0257788     0.74   0.459    -.0314294    .0696215
     Year_43 |   .0063293   .0245894     0.26   0.797    -.0418651    .0545237
     Year_44 |          0  (omitted)
       _cons |  -.0886041   .0302659    -2.93   0.003    -.1479241   -.0292841
-------------+----------------------------------------------------------------
     sigma_u |  .06242908
     sigma_e |  .10570513
         rho |  .25860259   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

4. Appendix. Stata Routine for panel regression

- Random effect model (RE) and the time effect
  - Fit RE model with time effect

. test Year_2 Year_3 Year_4 Year_5 Year_6 Year_7 Year_8 Year_9 Year_10 Year_11…Year_41 Year_42 Year_43 Year_44

```
(39)  Year_40 = 0
(40)  Year_41 = 0
(41)  Year_42 = 0
(42)  Year_43 = 0
(43)  o.Year_44 = 0
      Constraint 43 dropped

         chi2( 42) =  174.64
       Prob > chi2 =    0.0000
```

Time effect should be considered.

4. Appendix. Stata Routine for panel regression

- Which one is more proper?
  - Hausman test

. hausman xt_fe1 xt_re1

	Coefficients (b) xt_fe1	(B) xt_re1	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
rDEP	.1879095	.2169855	-.029076	.0178741
Year_2	.0109764	.0057601	.0052163	.0031059
Year_3	.0241612	.0190961	.0050651	.0030098
Year_4	.0955003	.0906024	.0048979	.0029034
Year_40	-.0011226	-.0009053	-.0002173	.
Year_41	.0070254	.0072429	-.0002176	.
Year_42	.0154131	.0155929	-.0001798	.
Year_43	.0062214	.0063293	-.0001079	.

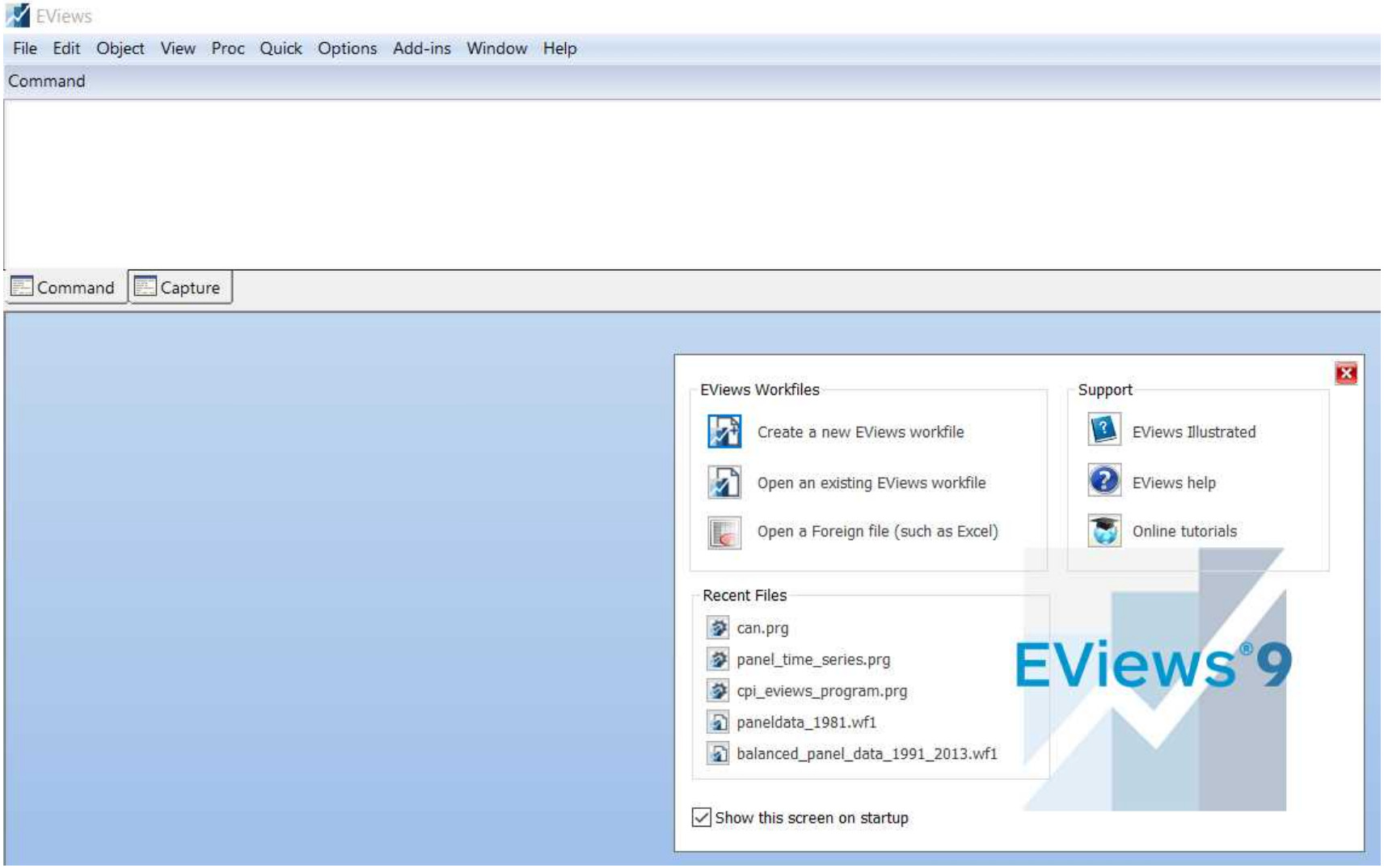
b = consistent under Ho and Ha; obtained from xtreg
B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

chi2(43) = (b-B)'[(V_b-V_B)^(-1)](b-B)
= 2.65
Prob>chi2 = 1.0000
(V_b-V_B is not positive definite)

Random effect model is better.

5. Appendix: Eviews



Prepare data: Tutorial01

T33 | f_x 4.601998495

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	code	i	t	complete_	count	RPI	CPI	GDP	GDPPW	TPOP	YNGPOP	WKGPOP	ELDPOP	rYNGPOP	rWKGPOP	rAGING
2	JPN	1	1981	1	44	126.2576	80.95833	54.86222	56.44489	91.86226	160.9576	97.19607	37.07345	0.2328	0.67453	0.09267
3	JPN	1	1982	1	44	132.5399	83.15833	56.7147	57.86576	92.4877	159.8205	98.01082	38.19314	0.229592	0.675585	0.094823
4	JPN	1	1983	1	44	136.3799	84.71667	58.45059	59.09713	93.12017	158.3005	98.90597	39.34176	0.225864	0.677124	0.097011
5	JPN	1	1984	1	44	137.6389	86.65833	61.05977	61.15786	93.71281	156.1878	99.83961	40.55303	0.22144	0.679194	0.099366
6	JPN	1	1985	1	44	138.1697	88.425	64.9269	64.38955	94.2875	153.4499	100.8345	41.87776	0.216233	0.681781	0.101986
7	JPN	1	1986	1	44	140.9285	88.95833	66.76503	65.50855	94.86375	150.0675	101.918	43.33618	0.210182	0.684921	0.104897
8	JPN	1	1987	1	44	151.2556	89.08333	69.50735	67.5141	95.33146	145.9665	102.9524	44.87441	0.203435	0.688478	0.108087
9	JPN	1	1988	1	44	159.0461	89.675	74.47483	71.65845	95.73905	141.4952	103.9303	46.52153	0.196364	0.692058	0.111578
10	JPN	1	1989	1	44	167.6261	91.71667	78.47424	74.86382	96.1318	137.0908	104.8227	48.30293	0.189474	0.695149	0.115377
11	JPN	1	1990	1	44	184.843	94.5	82.84714	78.51496	96.46053	132.9646	105.5176	50.19249	0.183145	0.697373	0.119482
12	JPN	1	1991	1	44	186.6228	97.61667	85.60126	80.73294	96.76037	129.2829	106.0302	52.20567	0.177522	0.698589	0.123889
13	JPN	1	1992	1	44	176.4367	99.28333	86.30236	81.15943	97.00086	125.9721	106.3368	54.31728	0.172547	0.698872	0.12858
14	JPN	1	1993	1	44	166.7786	100.5417	86.44999	81.15939	97.24057	123.0596	106.5188	56.5413	0.168142	0.698342	0.133515
15	JPN	1	1994	1	44	161.7864	101.2333	87.19967	81.71724	97.57242	120.5436	106.709	58.91414	0.164144	0.69721	0.138645
16	JPN	1	1995	1	44	159.4041	101.1083	88.89339	83.1744	97.94565	118.2655	106.8759	61.39445	0.160429	0.69564	0.143932
17	JPN	1	1996	1	44	156.2452	101.2417	91.21355	85.37636	98.19396	116.0242	106.837	63.88068	0.15699	0.693628	0.149382
18	JPN	1	1997	1	44	151.3936	103.025	92.66898	86.84327	98.4282	113.9926	106.7083	66.43363	0.153874	0.691144	0.154982
19	JPN	1	1998	1	44	147.863	103.7083	90.81269	85.22464	98.69603	112.211	106.5568	69.05138	0.151058	0.68829	0.160652
20	JPN	1	1999	1	44	143.7015	103.3667	90.63166	85.28309	98.8764	110.5279	106.2716	71.60284	0.148521	0.685195	0.166284
21	JPN	1	2000	1	44	139.2041	102.6917	92.67767	87.4765	99.04193	109.015	105.9458	74.10357	0.146243	0.681953	0.171804
22	JPN	1	2001	1	44	134.1572	101.8667	93.0071	88.0055	99.28086	107.7593	105.6833	76.59957	0.144211	0.678626	0.177163
23	JPN	1	2002	1	44	128.6769	100.5313	93.2764	88.5018	99.51199	106.6574	105.3949	79.04432	0.142405	0.675202	0.182393
24	JPN	1	2003	1	44	120.519	100.7	94.84822	90.28103	99.72515	105.678	105.0589	81.47233	0.140796	0.671611	0.187594
25	JPN	1	2004	1	44	113.1783	100.6917	97.08733	92.91729	99.75873	104.6276	104.4879	83.81155	0.139349	0.667736	0.192915
26	JPN	1	2005	1	44	107.951	100.4167	98.35211	94.72023	99.7681	103.6508	103.8343	86.2321	0.138035	0.663497	0.198468
27	JPN	1	2006	1	44	104.4856	100.6583	100.0171	96.93157	99.83134	102.816	103.1832	88.79831	0.136837	0.658919	0.204244
28	JPN	1	2007	1	44	103.3809	100.7167	102.2097	99.67909	99.94612	102.1096	102.5387	91.4962	0.135741	0.654051	0.210208
29	JPN	1	2008	1	44	100.3465	102.1	101.145	99.37993	99.99453	101.3954	101.7761	94.23763	0.134726	0.648872	0.216402
30	JPN	1	2009	1	44	97.86816	100.725	95.55476	94.70398	99.98204	100.6652	100.8984	97.0427	0.133772	0.643357	0.222871
31	JPN	1	2010	1	44	100	100	100	100	100	100	100	100	0.132864	0.637514	0.229622
32	JPN	1	2011	1	44	100.3488	99.71667	99.54728	100.7214	99.80267	99.14327	98.83433	102.8727	0.131987	0.631329	0.236685

Program:Setup

- ' ----------'
  Data Load, Setup
- ' ----------

- ' Set working directory (to a designated folder)
- cd “E:¥EViews7¥panel_time_series"

Import data and set panel data

- ' Opening csv file as a workfile
- wfopen(page=data) "paneldata_10countries_1970_2013.csv"

- ' Set data structure as a balanced panel data
- pagestruct code @date(t)

Time series data

- ' ----------
- ' Generate ln series, dln series, d2ln series
- ' ----------

- ' Level in Natural Log (ln_x = log(x))
- series ln_hpi = log(hpi)
- series ln_gdppw = log(gdppw)
- series ln_tpop = log(tpop)
- series ln_olddep = log(rolddep)

- ‘show’ or click the variable name to see the data

Time series data: Difference data

- ' 1st Difference in Natural Log (dln_x = log(x) - log(x(-1)))
- series dln_hpi = dlog(hpi)
- series dln_gdppw = dlog(gdppw)
- series dln_tpop = dlog(tpop)
- series dln_olddep = dlog(rolddep)

- ' 2nd Difference in Natural Log (d2ln_x = dln_x - dln_x(-1))
- series d2ln_hpi = dlog(hpi, 2)
- series d2ln_gdppw = dlog(gdppw, 2)
- series d2ln_tpop = dlog(tpop, 2)
- series d2ln_olddep = dlog(rolddep, 2)

Histogram and plot

- ' Set sample in 2000
- smpl 2000 2000
- ' Histogram and Descriptive Statistics
- ln_hpi.hist
- ln_gdppw.hist
- ln_tpop.hist
- ln_olddep.hist

- ' Scatterplot: set up group>draw the scat
- group group_hpi_olddep ln_hpi ln_olddep
- group_hpi_olddep.scat

Summary Stats

- ' Reset sample in 1970-2013
- smpl @all (or just smpl 1970 2013)

- ' Summary Stats
- hpi.stats
- gdppw.stats
- tpop.stats
- rolddep.stats

Create and save graphs as "g_..."

- ' Graph of ln_x

- graph g_ln_hpi.line(panel=individual) ln_hpi
- show g_ln_hpi
- graph g_ln_gdppw.line(panel=individual) ln_gdppw
- show g_ln_gdppw
- graph g_ln_tpop.line(panel=individual) ln_tpop
- show g_ln_tpop
- graph g_ln_olddep.line(panel=individual) ln_olddep
- show g_ln_olddep

- ' Graph of dln_x

- graph g_dln_hpi.line(panel=individual) dln_hpi
- show g_dln_hpi
- graph g_dln_gdppw.line(panel=individual) dln_gdppw
- show g_dln_gdppw
- graph g_dln_tpop.line(panel=individual) dln_tpop
- show g_dln_tpop
- graph g_dln_olddep.line(panel=individual) dln_olddep
- show g_dln_olddep

- ' Multiple graph of ln_x1 and ln_x2

- graph g_ln_hpi_olddep.line(panel=individual) group_hpi_olddep

- ' set second axis to make it readable
- g_ln_hpi_olddep.setelem(2) axis(right)
- show g_ln_hpi_olddep

Run a multiple regression model

- ' Regression
- equation reg01.ls dln_hpi c dln_gdppw dln_olddep dln_tpop
- show reg01

- ' Saving workfile

- wfsave “paneldata_10countries_1970_2013.wf1”

清水千弘 : Chihiro Shimizu, PhD

シンガポール国立大学不動産研究センター　教授

Professor, Institute of Real Estate Studies
National University of Singapore
21 Heng Mui Keng Terrace, #04-02
Singapore 119613
Tel: (65) 6601 4925 Fax: (65) 6774 1003
Email: cshimizu@nus.edu.sg
Website: www.ires.nus.edu.sg