RM01 Research Methods

SPSS Exercises

SPSS Exercises Solutions

(Version: November 2021)

By Helen Bao

Department of Land Economy

University of Cambridge

SPSS Exercises

- 1. Quarterly data on property price index and GDP are given in Worksheet 'example1'. Variable T is a quarterly time index (e.g., T = 1 for the first quarter). Use SPSS to complete the following tasks.
 - a) Open the file in SPSS and view the data

File Edit View	v Data Transform	Analyze	Graphs Uti		
New	,				
Open		🔒 Dat	ta		
Import Data	>	- 🗖 Inte	ernet Data		
	921	📈 🛅 Syr	ntax		
	00	🔪 🔚 Ou	tput		
ave Save	ж:		int		
Save As					
Step 2:					
		Downloads	\$	Q Search	
Favourites	Name	_	 Date Modified 	Date Created	
Dropbox (Per	SPSS Exercises - Data	.xls 🤳	Today at 13:40	Today at 13:40	
A Applications	SPSS Exercises - Que	stions.pdf	Today at 13:40	Today at 13:40	
Applications	SPSS Exercises - Solu	tions.pdf	Today at 13:40	Today at 13:40	
U Downloads	SPSS installation instr	uctions.pdf	Today at 13:40	Today at 13:40	
🛅 LUP	STATA Exercises - Dat	a.xlsx	Today at 13:30	Today at 13:30	
Claud	STATA Exercises - Que	estions.pdf	Today at 13:30	Today at 13:30	
	🛃 STATA Exercises - Sol	utions.pdf	Today at 13:30	Today at 13:30	
Desktop	STATA Exercises - STA	TA codes.do	Today at 13:39	Today at 13:29	
🖺 Documents					
iCloud Drive					
Locations					
🗐 Macintosh HD					
Remote Disc					
Network					

Note: You may need to change the 'Files of types' to be 'Excel' in order to make Excel files visible in this window.

Step 3:

Range:								
Read variable names from first row of data								
🗹 Percent	age of value	s that deter	mine data t	ype:	95			
Ignore l	hidden rows	and columr	ıs					
Remove	leading sp	aces from st	ring values					
	. icauing spa							
Remove trailing spaces from string values								
Preview								
	🔗 INDEX	🧳 GDP	🛷 т					
	V							
1	101	28456500	1					
1 2	101 101	28456500 18401000	1					
1 2 3	101 101 103	28456500 18401000 10854284	1 1 1					
1 2 3 4	101 101 103 99	28456500 18401000 10854284 3863400	1 1 1 1			1		
1 2 3 4 5	101 101 103 99 100	28456500 18401000 10854284 3863400 2111188	1 1 1 1 1					
1 2 3 4 5 6	101 101 103 99 100 102	28456500 18401000 10854284 3863400 2111188 12364727	1 1 1 1 1 1					
1 2 3 4 5 6 7	101 101 103 99 100 102 97	28456500 18401000 10854284 3863400 2111188 12364727 12356400	1 1 1 1 1 1 1 1			(

Data view: to view and edit data

	💉 INDEX	🧳 GDP	💑 т	var	var		Τ
1	101.0	28456500	1				
2	100.7	18401000	1				
3	103.1	10854284	1				
4	99.1	3863400	1				
5	100.3	2111188	1				
6	101.9	12364727	1				•
7	96.6	12356400	1				
8	100.1	10030125	1				
9	102.5	11201156	1				
10	105.1	49508400	1				
11	100.2	11503034	1				
						_	



Variable view: to view and edit variable attributes

	Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	INDEX	Numeric	8	1		None	None	11	🗮 Right	🛷 Scale
2	GDP	Numeric	11	0		None	None	11	🗮 Right	🛷 Scale
3	Т	Numeric	11	0		None	None	11	🗮 Right	뤚 Nominal
4										
5										
6										
	Data View Variable View									

b) Create a histogram of variable INDEX

Step 1	:								
File	Edit	View	Data	Transform	Analyze	Graphs	Utilities	Extensions	Window
					Untitled	👘 CI	hart Builder.		-
	r			•=	P H	+ Co	aphboard 1 ompare Sub egression Va	emplate Choo groups ariable Plots	oser
						+ W	eibull Plot		
INDEX 101	0	GDP 2845650	00	♣ T 1	var	Lega	cy Dialogs		
Step 2	:								
Variables	::	C	Chart previe	ew uses example data			Element Properties	Chart Appearance	Options
No cate	x egories (scale ariable)	Histogram	Simple F	listogram of INDE	×	Barl X-Ax Y-Ax Tritle Stat Var Stat	isi (Bar1) s1 (Bar1) 1 istics able: INDEX istic: istic: istogram Display normal curv Display error bars r Bars Represent Confidence interva	Set Parat	eters
	Gallery	Basic Elemer	its Grou	ips/Point ID Titles/	Footnotes		Level (%): 95	15	
Choo Favor Bar Line Area Pie/P Kisto Histo Hig Boxpi Dual	se from: ites olar ec/Dot gram lov Axes					Bar S	Standard error Multiplier: 2 Standard deviation Multiplier: 2 tyle: Bar		•
	?	Reset	Pas	te Cancel	ОК				





c) Create a scatter plot between INDEX and GDP

Step 1:							
File Edit	View Da	ta Transform	Analyze	Graphs	-Utilities	Extensions	Windo
			Untitled	📊 Cha	rt Builder.		
	∽∎	▮ !		E Gra	phbeard T npare Sub ression Va bull Plot	emplate Choo groups rriable Plots	ser
INDEX 🔬	GDP	- т	var		built lot		
101.0	28456500	1		Legacy	/ Dialogs		
Step 2: Variables:	Chart p	review uses example data	2	Edit Prope Point1	ent Properties	Chart Appearance O	ptions
CDP T T No categories (scale variable) Choose from: Favorites Bar Line Area Pie/Polar Scatter/Dot Histogram High-Low Boxplot Dual Axes	Asic Elements	GDP Groups/Point ID Titles/Fe	Set color?	X-Axis1 (I Y-Axis1 (I Title 1 Axis Labe Scale Ra Variable Minimu Maxim Major I Origin Scale Ty Typ	Point1) Point1) Point1) el: [NDEX nge IMDEX Autom Im	atic Custom 0 0 0 0 0 0 0 0 0 0 0 0 0	
?	Reset	Paste Cancel	ОК				





d) Generate descriptive statistics for INDEX and GDP

Step 1:									
File E	dit View	Data	Transform	Analyze	Graphs	Window	Help		
				Power A	Analysis	ata Editor			
			▙ੂ	Reports	6	Q			
• • •				Descrip	tive Statis	123 Freq	uencies		
1				Bayesia	n Statistic	CHa Desc	criptives		
-				Tables			•	A Expl	ore
INDEX	🧳 GDP		💑 т	Compar	re Means		•	Cros	stabs
101.0	284565	500	1	Genera	l Linear Mo	odel	•	TUR	F Analysis
100.7	184010	000	1	Genera	lized Linea	r Models	•	🔽 Ratio	D
103.1	108542	284	1	Mixed N	Nodels			P-P	Plots
99.1	38634	100	1	Correla	te sion			🛃 Q-Q	Plots



Step 3:



Output:

					Descriptive	Statistics							
	N	Range	Minimum	Maximum	Sum	Me	an	Std. Deviation	Variance	Skev	vness	Kur	tosis
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
NDEX	100	32.6	96.5	129.1	10326.0	103.260	.4912	4.9119	24.127	2.493	.241	8.277	.478
GDP	100	61463225	1044875	62508100	1.E+9	13067838.9	1090432.76	10904327.6	1.189E+14	2.061	.241	5.885	.478
Valid N (listwise)	100												

e) Generate frequency statistics for T

Step 1:										
File Edit	View	Data	Transform	Analyze	Graphs	Utilities	Extension	s Wir	ndow	Help
				Power A	Analysis		1	► ata E	ditor	
				Reports Descrip Bayesia Tables	<mark>tive Statis</mark> n Statistic	tics s	1		Freq Dest	uencies cri pt iv es ore
INDEX	GDP 284565 184010 108542 38634	00 00 84 00	T 1 1 1 1 1	Compar General General Mixed N Correla	e Means Linear Mo lized Linea Iodels te	odel nr Models			E Cros TUR Ratic P-P Q-Q	stabs F Analysis o Plots Plots
Step 2:										
			Freque	ncies		_				
 ✓ INDEX ✓ GDP ✓ Display fi ? 	requency Rest	tables	Variab T Ci Paste	reate APA st	yle tables Cance	Statisti Chart: Forma Style Bootstr	cs s tt ap DK			
	Freque	ncies: (Charts							
 None Bar char Pie char Histogr Show 	rts rts rams: w normal	curve	on histogram							
Chart Values	s ncies C	Perce Cancel	ntages	nue	2					
							Ŧ			

Output:

			т		
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	35	35.0	35.0	35.0
	2	35	35.0	35.0	70.0
	3	30	30.0	30.0	100.0
	Total	100	100.0	100.0	



f) Generate two new variables: $T2 = T^*T$, and LNGDP = ln(GDP)

Step 1:

File E	dit	View	Data	Transform	Analyze	Graphs	Utilities		
				📄 Compu	ıte Variable				
1		3		 Programmability Transformation Count Values within Cases Shift Values 					
INDEX		🔗 GDP		Recode	e into Same	e Variables	S		
101.0		284565	500		atic Recodu	rent variai	oles		
100.7		184010	000	+ Create	Dummy Va	ariables			

Step 2:

	Compute Variable	
Targer Variable: T2 Type & Label	Compute Variable Numeric Expression: T ** 2 + < > 7 8 9 - <= >= 4 5 6 * = ~= 1 2 3 / & 0 .	Function group: All Arithmetic CDF & Noncentral CDF Conversion Current Date/Time Date Arithmetic Date Creation
If (optional ca ? Reset	<pre>se selection condition)</pre>	Functions and Special Variables: \$Casenum \$Date \$Date11 \$JDate \$Sysmis \$Time Abs Any Applymodel Arsin Artan Cancel OK

8

Step 3:

File E	dit	View	Data	Transform	Analyze	Graphs	Utilities I
				📄 Compu	ıte Variable		
1		3		🛨 Pogra 🕜 Count Shift Value	mmability Values with es	Transform nin Cases.	ation
INDEX		🔗 gdp		Recode	e into Same	e Variables	S
101.0		28456	500	Recode	e Into Diffe	rent variai	oles
100.7		184010	000	+ Create	Dummy Va	ariables	-

Step 4: Target Variable: LNGDP Type & Label M INDEX GDP	<pre> Numeric Explession: LN(GDP) </pre>	
₩ Т	+ > 7 8 9 - <=	All Arithmetic CDF & Noncentral CDF Conversion Current Date/Time Date Arithmetic Date Creation Functions and Special Variables: Abs Arsin Artan Cos Exp Lg10 Ln
If (optional ca	ase selection condition)	Mod Rnd(1) Rnd(2)
? Reset	Paste	Cancel

g) Estimate the regression model $INDEX = \beta_0 + \beta_1 GDP + \beta_2 T + \beta_3 T2 + \varepsilon$, where T2 is T squared. Obtain collinearity statistics and autocorrelation test statistics

Step	1:											
File	Edit	View	Data	Transform	Analyze	Graphs	Utilities	Extensions	Window	Help		
					Power /	Analysis		•	ata Editor			
1.00					Reports Descrip Bayesia Tables	s otive Statis on Statistic	atics cs	* *	Q			
NDEX		🔗 GDP		💑 т 👘	Compa	re Means		•	va	r	var	var
101.0	D	28456	500	1	Genera	l Linear M	odel	•				
100.7	7	184010	000	1	Genera	lized Linea	ar Models	•				
103.1	1	108542	284	1	Mixed N Correla	Aodels		·				
99.1	1	38634	400	1	Regress	sion			Auto	omatic	Linear Mo	delina
100.3	3	2111	188	1	Logline	ar		•	R Line	ar	>	5
101.9	Э	12364	727	1	Neural	Networks			Cur	ve Estin	nation	
96.6	5	123564	400	1	Classify	ion Peduc	tion		🔣 Part	ial Leas	st Squares	5





Step 3:

🛑 🔵 🛑 🛛 Linear Regr	ession: Statistics
Regression Coefficients Estimates Confidence intervals Level(%): 95 Covariance matrix Residuals Uurbin-Watson	 Model fit R squared change Descriptives Part and partial correlations Collinearity diagnostics
 Casewise diagnostics Outliers outside: All cases 	3 standard deviations
?	Cancel

Output:

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin- Watson
1	.439 ^a	.192	.167	4.4827	1.596

a. Predictors: (Constant), T2, GDP, T

b. Dependent Variable: INDEX

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	459.489	3	153.163	7.622	.000 ^b
	Residual	1929.071	96	20.094		
	Total	2388.560	99			

a. Dependent Variable: INDEX

b. Predictors: (Constant), T2, GDP, T

Coefficients^a

		Unstandardized Coefficients		Standardized Coefficients			Collinearity	Statistics
Model		В	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	99.780	3.356		29.728	.000		
	GDP	1.391E-7	.000	.309	3.301	.001	.961	1.040
	Т	078	3.780	013	021	.984	.022	46.049
	Т2	.408	.942	.270	.433	.666	.022	46.157

h) Use stepwise selection method to determine the best set of regressors to predict the value of INDEX

Include all variables in the 'independent(s)' list. Choose model selection methods (e.g., backward or forward) from the 'Method..' dropdown menu.

GDP	Dependent:	Statistics
	RIOCK 1 OF 1	Piots
🖗 LNGDP	Previous	Save
	Independent(s):	Options
	LNGDP	Style
	GDP Enter Method: ✓ Stepwise Remove Backward Forward	Bootstrap
	Case Labels: WLS Weight:	
? Reset	Paste	Cancel OK

Outputs (Stepwise selection):

Model Summary^c

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin- Watson
1	.359 ^a	.129	.120	4.6078	
2	.439 ^b	.192	.176	4.4595	1.596

a. Predictors: (Constant), GDP

b. Predictors: (Constant), GDP, T2

c. Dependent Variable: INDEX

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	307.798	1	307.798	14.497	.000 ^b
	Residual	2080.762	98	21.232		
	Total	2388.560	99			
2	Regression	459.480	2	229.740	11.552	.000 ^c
	Residual	1929.080	97	19.887		
	Total	2388.560	99			

a. Dependent Variable: INDEX

b. Predictors: (Constant), GDP

c. Predictors: (Constant), GDP, T2

Coefficients^a

		Unstandardize	d Coefficients	Standardized Coefficients			Collinearity	Statistics
Model		В	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	101.147	.721		140.220	.000		
	GDP	1.617E-7	.000	.359	3.807	.000	1.000	1.000
2	(Constant)	99.713	.870		114.608	.000		
	GDP	1.391E-7	.000	.309	3.320	.001	.962	1.040
	Т2	.388	.141	.257	2.762	.007	.962	1.040

i) Generate a scatter plot between the residuals and GDP for the regression model *INDEX* = $\beta_0 + \beta_1 GDP + \beta_2 T + \varepsilon$.

Step 1: Choose GDP and T only as the independent variables, and use 'Enter' method. Click the 'Save' button.

	Linear Regression
 ✓ GDP ♣ T ♣ T2 ✓ LNGDP 	Dependent: NDEX Block 1 of 1 Pevious Independent(s): CDP T Method: Enter Selection Variable: Statistics Statistics Statistics Statistics Statistics Statistics Options Bootstrap
? Re	Rule Case Labels: WLS Weight: Set Paste Cancel OK

Step 2: Check the 'Unstandardized' checkbox below 'Residuals'. Click 'Continue'.

Linear Regress	ion: Save
Predicted Values	Residuals
Unstandardized	🗹 Unstandardized
Standardized	🗌 Standardized 🥜
Adjusted	Studentized
S.E. of mean predictions	Deleted
	Studentized deleted
Distances	Influence Statistics
Mahalanobis	DfBetas
Cook's	Standardized DfBetas
Leverage values	DfFits
Prediction Intervals	Standardized DfFits
🗌 Mean 📄 Individual	Covariance ratios
Confidence Interval: 95 %	
Coefficient statistics	
Create coefficient statistics	
• Create a new dataset	
Dataset name:	
🔵 Write a new data file	
File	
Export model information to XML file	
	Browse
✓ Include the covariance matrix	
?	Cancel

A new variable 'RES_1' is created to store the residuals of your regression model.

24 : RES_1	-1.704	477237491771				<u> </u>
	💉 INDEX	🧳 GDP	💦 Т	뤚 Т2	🛷 LNGDP	🥟 RES_1
1	101.0	28456500	1	1.00	17.16	-2.95386
2	100.7	18401000	1	1.00	16.73	-1.84494
3	103.1	10854284	1	1.00	16.20	1.61246
4	99.1	3863400	1	1.00	15.17	-1.40802
5	100.3	2111188	1	1.00	14.56	.03749
6	101.9	12364727	1	1.00	16.33	.20082
7	96.6	12356400	1	1.00	16.33	-5.09801
8	100.1	10030125	1	1.00	16.12	-1.27207
9	102.5	11201156	1	1.00	16.23	.96386
10	105.1	49508400	1	1.00	17.72	-1.80353
11	100.2	11503034	1	1.00	16.26	-1.37844
12	104.3	15680138	1	1.00	16.57	2.13629
13	109.0	13126854	1	1.00	16.39	7.19404
14	100.4	3634412	1	1.00	15.11	07594
15	102.8	10742289	1	1.00	16.19	1.32815
16	103.9	5583268	1	1.00	15.54	3.15100

Step 3: Follow the instructions in part c).





Scatter Plot of Unstandardized Residual by GDP



j) Perform a White heteroskedasticity test on the model *INDEX* = $\beta_0 + \beta_1 GDP + \beta_2 T + \varepsilon$.

	Linear Regression		🛑 🔘 🛑 🛛 Linear Regress	sion: Save
 	Dependent: VINDEX Block 1 of 1 Previous Next Independent(s): CDP T	Statistics Plots Save Options Style	Predicted Values Unstandardized Standardized Adjusted S.E. of mean predictions	Residuals Vinstandardized Standardized Studentized Deleted Studentized deleted
	Method: Enter > Selection Variable: Case Labels: WLS Weight:	Bootstrap	Distances Mahalanobis Cook's Leverage values Prediction Intervals Mean Individual Confidence Interval: 95 % Coefficient statistics	Influence Statistics DfBetas Standardized DfBetas DfFits Standardized DfFits Covariance ratios
? Reset	Paste	cel OK	Create a new dataset	

Step 1: Request the unstandardized residuals to be saved as a new variable.



Target Variable:		Numeric Expression:	
RESI2	=	RES_1 **2	
Type & Label			
	•		
			Function group:
		+ $<$ $>$ 7 8 9	All
			Arithmetic
		- <= >= 4 5 6	CDF & Noncentral CDF
V Unstandardized K			Conversion

Step 3: Create the squared and cross terms of all independent variables, i.e., GDP2 = GDP*GDP, GDPT = GDP*T. Note that T2 has already been created in Part f).

	🔗 INDEX	🧳 GDP	💦 Т	💦 Т2	🧳 LNGDP	NES_1	🛷 RESI2	🔗 GDP2	🧳 GDPT
1	101.0	28456500	1	1.00	17.16	-2.95386	8.73	8.10E+14	28456500.00
2	100.7	18401000	1	1.00	16.73	-1.84494	3.40	3.39E+14	18401000.00
3	103.1	10854284	1	1.00	16.20	1.61246	2.60	1.18E+14	10854284.00
4	99.1	3863400	1	1.00	15.17	-1.40802	1.98	1.49E+13	3863400.00
5	100.3	2111188	1	1.00	14.56	.03749	.00	4.46E+12	2111188.00
6	101.9	12364727	1	1.00	16.33	.20082	.04	1.53E+14	12364727.00
7	96.6	12356400	1	1.00	16.33	-5.09801	25.99	1.53E+14	12356400.00
8	100.1	10030125	1	1.00	16.12	-1.27207	1.62	1.01E+14	10030125.00
9	102.5	11201156	1	1.00	16.23	.96386	.93	1.25E+14	11201156.00
10	105.1	49508400	1	1.00	17.72	-1.80353	3.25	2.45E+15	49508400.00
11	100.2	11503034	1	1.00	16.26	-1.37844	1.90	1.32E+14	11503034.00
12	104.3	15680138	1	1.00	16.57	2.13629	4.56	2.46E+14	15680138.00
13	109.0	13126854	1	1.00	16.39	7.19404	51.75	1.72E+14	13126854.00
14	100.4	3634412	1	1.00	15.11	07594	.01	1.32E+13	3634412.00

Step 4: Estimate the following regression model

$$RESI2 = \beta_0 + \beta_1 GDP + \beta_2 T + \beta_3 GDP2 + \beta_4 T2 + \beta_5 GDPT + \varepsilon$$



Output:

			ANOVA ^a			
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	75741.626	5	15148.32	12.320	<.001 ^b
	Residual	115584.250	94	1229.620		
	Total	191325.876	99			
- D	anandant Vari					

a. Dependent Variable: RESID2b. Predictors: (Constant), GDPT, T, GDP2, GDP, T2

Coefficients^a

		Unstandardize	d Coefficients	Standardized Coefficients		
Model		В	Std. Error	Beta	t	Sig.
1	(Constant)	12.938	27.508		.470	.639
	GDP	-3.337E-6	.000	828	-3.087	.003
	Т	9.694	29.608	.178	.327	.744
	Т2	-5.427	7.453	401	728	.468
	GDP2	2.527E-14	.000	.327	1.494	.139
	GDPT	1.767E-6	.000	1.152	4.143	<.001
a. D	ependent Va	riable: RESID2			_	

Because the F test statistic is significant at the 5% level, the model suffers from heteroskedasticity problem. This may be caused by the interaction term between GDP and T, as the coefficient estimate of GDPT is significant.

k) Perform a RESET test on the final model from part h)

Step 1	: Request the	unstandardised	predicted	value of I	NDEX	to be s	saved as	a new	variable.
--------	---------------	----------------	-----------	------------	------	---------	----------	-------	-----------



Step 2: Generate the squared and cubed terms for the unstandardised predicted value of INDEX. Name them as INDEX2 and INDEX3 respectively.

PRE_1	💉 INDEX2	💉 INDEX3
103.95386	10806.41	1123367.58
102.54494	10515.47	1078307.83
101.48754	10299.72	1045293.39
100.50802	10101.86	1015318.17
100.26251	10052.57	1007896.01
101.69918	10342.72	1051846.37
101.69801	10342.49	1051810.17
101.37207	10276.30	1041729.33
101.53614	10309.59	1046795.87
106.90353	11428.37	1221732.59
101 57844	10318 18	1048104 62

Step 3: Estimate the full model by including INDEX2 and INDEX3 as new regressors.



Full model output: Note that INDEX2 is excluded from the model due to collinearity. Therefore, the full model is $INDEX = \beta_0 + \beta_1 GDP + \beta_2 T + \beta_3 INDEX3 + \varepsilon$.

			ANOVA"			
Mode	I	Sum of Squares	df	Mean Square	F	Sig.
1	Regression	690.092	3	230.031	13.002	.000 ^b
	Residual	1698.468	96	17.692		
	Total	2388.560	99			

The reduced model is $INDEX = \beta_0 + \beta_1 GDP + \beta_2 T + \varepsilon$. The output is

ANOVA ^a							
Model		Sum of Squares	df	Mean Square	F	Sig.	
1	Regression	455.728	2	227.864	11.435	.000 ^b	
	Residual	1932.832	97	19.926			
	Total	2388.560	99				

$$F = \frac{(1932.832 - 1698.468)/1}{1698.468/96} = \frac{234.364}{17.692} = 13.2466 > FINV(0.05,1,96) = 3.9402$$

Reject the null hypothesis. The model has misspecification problems.

l) Perform a predictive failure test on the model *INDEX* = $\beta_0 + \beta_1 GDP + \beta_2 T + \varepsilon$. Reserve the last 10 observations for the test.

Step 1: Create a new variable PRED, which equals 1 for the last 10 observations only, and zero otherwise.



Outputs:

		🛷 RESI2	🛷 GDP2	🧳 GDPT	🛷 RES_2	PRE_1	🧳 INDEX2	INDEX3	RED
84	82475	.68	2.35E+13	1.14E+20	-9.78644	103.72475	10758.82	1115956.17	.00
85	-3.93353	15.47	1.82E+14	2.45E+21	-6.68887	104.93353	11011.05	1155427.89	.00
86	10968	.01	5.77E+13	4.38E+20	-14.13426	104.10968	10838.83	1128426.75	.00
87	2.15659	4.65	4.11E+13	2.63E+20	-7.91759	103.94341	10804.23	1123028.82	.00
88	.24174	.06	1.86E+14	2.55E+21	-22.35863	104.95826	11016.24	1156244.99	.00
89	12.06013	145.45	3.17E+14	5.64E+21	116.53470	105.53987	11138.66	1175573.02	.00
90	-2.28969	5.24	1.22E+14	1.34E+21	-13.48424	104.58969	10939.00	1144107.0	.00
91	87419	.76	2.76E+14	4.59E+21	-26.18894	105.37419	11103.72	1170045.6 <mark>4</mark>	1.00
92	-3.14759	9.91	8.64E+13	8.03E+20	-6.49371	104.34759	10888.42	1136180.2 <mark>2</mark>	1.00
93	-10.14487	102.92	1.22E+15	4.28E+22	27.24331	107.94487	11652.09	1257783.79	1.00
94	-3.60212	12.98	8.38E+14	2.43E+22	-41.35644	107.10212	11470.86	1228553.88	1.00
95	-2.04936	4.20	2.53E+13	1.27E+20	-6.50566	103.74936	10763.93	1116750.7	1.00
96	.93409	.87	5.24E+12	1.20E+19	-6.00757	103.36591	10684.51	1104414.23	1.00
97	4.70140	22.10	5.07E+14	1.14E+22	-15.76819	106.19860	11278.14	1197722.91	1.00
98	.23349	.05	3.50E+14	6.55E+21	-30.42572	105.66651	11165.41	1179809.96	1.00
99	-2.07897	4.32	1.45E+13	5.53E+19	-4.71333	103.57897	10728.60	1111257.54	1.00
100	-3.88299	15.08	6.59E+13	5.35E+20	.23754	104.18299	10854.10	1130812.08	1.00

Step 2: Estimate the reduced model (*INDEX* = $\beta_0 + \beta_1 GDP + \beta_2 T + \varepsilon$) by using the first 90 observations only.

	L11001 110910001011		
	Dependent:		.00
GDP		Statistics	.00
🕹 T	Block 1 of 1	Plots	.00
💑 Т2			.00
V LNGDP	Previous	Save	1.00
Providence Research	Independent(s):	Options	1.00
GDP2	GDP	Style	1.00
🛷 GDPT	● [●] ¹	Style	1.00
🧳 Unstandardized R		🛑 🔘 🔵 🛛 Linear Reg	gression: Set Rule
Unstandardized Pr		Define Selection Rule	2
VINDEX2			P <u>RE</u> D
	Method: Enter 🗘	/	Value:
		equal to	0
	Selection Variable:		
	PRED=? Rule	?	Cancel Continue
	Case Labels:		

Model 1 Outputs (using the first 90 observations):

	ANOVA ^{a,b}									
Model		Sum of Squares	df	Mean Square	F	Sig.				
1	Regression	528.676	2	264.338	13.190	.000 ^c				
	Residual	1743.508	87	20.040						
	Total	2272.185	89							

Model 2 (*INDEX* = $\beta_0 + \beta_1 GDP + \beta_2 T + \varepsilon$) outputs (using all 100 observations): ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	455.728	2	227.864	11.435	.000 ^b
	Residual	1932.832	97	19.926		
	Total	2388.560	99			

$$F = \frac{(1932.832 - 1743.508)/10}{1743.508/87} = \frac{18.932}{20.040} = 0.9447 < FINV(0.05, 10, 87) = 1.9413$$

Do not reject the null hypothesis. The model predicts well.

m) Test if there is a structural break at GDP = 30,000,000.

Step 1: Create a dummy variable BREAK, which equals one when GDP > 30,000,000, and zero otherwise.

Step 2: Create interaction terms between BREAK and the two regressors GDP and T. Name them as GDPBREAK and TBREAK respectively.

	🧳 INDEX	🧳 GDP	💑 т	💑 BREAK	🧳 GDPBREAK	💑 TBREAK
34	104.7	1048208	1	.00	.00	.00
35	100.7	3150000	1	.00	.00	.00
36	99.0	32127100	2	1.00	32127100.00	2.00
37	101.6	20511600	2	.00	.00	.00
38	100.1	11868132	2	.00	.00	.00
39	103.4	4328880	2	.00	.00	.00
40	101.7	3167000	2	.00	.00	.00
41	99.9	14000238	2	.00	.00	.00
42	99.9	14061000	2	.00	.00	.00
43	96.5	11501802	2	.00	.00	.00
44	99.8	12321276	2	.00	.00	.00
45	109.7	54087600	2	1.00	54087600.00	2.00
46	102.5	12975725	2	.00	.00	.00
47	106.6	17818302	2	.00	.00	.00

Outputs:

Step 3: Estimate the full model *INDEX* = $\beta_0 + \beta_1 GDP + \beta_2 T + \beta_3 BREAK + \beta_4 GDPBREAK + \beta_5 TBREAK + \varepsilon$.

GDP	Dependent:	Statistics
BREAK	Block 1 of 1	Plots
	Previous Next	Save
	Independent(s): GDP	Options
-		Style
		Bootstrap
	• I DREAK	
-	Method: Enter ᅌ	

Full model outputs:

			ANOVA ^a			
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	909.611	5	181.922	11.563	.000 ^b
	Residual	1478.949	94	15.733		
	Total	2388.560	99			

a. Dependent Variable: INDEX

b. Predictors: (Constant), TBREAK, T, GDP, GDPBREAK, BREAK

Coefficients^a Standardized Coefficients Unstandardized Coefficients В Std. Error Beta Sig. Model t 1 (Constant) 98.877 1.174 84.205 .000 GDP .000 1.672 .098 1.045E-7 .232 т .262 1.588 .519 3.062 .003 -34.953 7.710 BREAK -1.825 -4.534 .000 **GDPBREAK** 6.940E-7 .000 1.612 4.825 .000 TBREAK 1.786 2.204 .224 .420 .810

Reduced model (*INDEX* = $\beta_0 + \beta_1 GDP + \beta_2 T + \varepsilon$) outputs:

ANOVA								
Model		Sum of Squares	df	Mean Square	F	Sig.		
1	Regression	455.728	2	227.864	11.435	.000 ^b		
	Residual	1932.832	97	19.926				
	Total	2388.560	99					

$$F = \frac{(1932.832 - 1478.949)/3}{1478.949/94} = \frac{151.294}{15.733} = 9.616 > FINV(0.05,3,94) = 2.7014.$$

Reject the null hypothesis. There is a structure break at GDP = 30,000,000.

n) Create dummy variables for T

Outputs:

	🧳 INDEX	🧳 GDP	🗞 т	💑 TD_1	💑 TD_2	💑 TD_3
32	101.9	3487465	1	1.00	.00	.00
33	100.4	1044875	1	1.00	.00	.00
34	104.7	1048208	1	1.00	.00	.00
35	100.7	3150000	1	1.00	.00	.00
36	99.0	32127100	2	.00	1.00	.00
37	101.6	20511600	2	.00	1.00	.00
38	100.1	11868132	2	.00	1.00	.00
39	103.4	4328880	2	.00	1.00	.00
40	101.7	3167000	2	.00	1.00	.00
41	99.9	14000238	2	.00	1.00	.00
42	99.9	14061000	2	.00	1.00	.00
43	96.5	11501802	2	.00	1.00	.00

o) Estimate a regression model using the group of dummy variables created in part n)

Outputs:

Coefficients^a

		Unstandardize	d Coefficients	Standardized Coefficients		
Model		В	Std. Error	Beta	t	Sig.
1	(Constant)	103.215	1.066		96.828	.000
	GDP	1.391E-7	.000	.309	3.301	.001
	T=1.0	-3.106	1.136	303	-2.733	.007
	T=2.0	-1.960	1.127	191	-1.740	.085

a. Dependent Variable: INDEX

p) Create a dummy variable INDEXD, which equals one when INDEX > 103 and zero otherwise

			Compute		
Farget Varia	ıble:	Numeric Expr	ression:		
INDEXD	=	0	•		
Type & L	abel	T			
🖋 INDEX					
🧳 GDP					
💑 T 💂 T – 1 0 ["		+ <	> 7		
T=2.0 [TD_2]				
💑 T=3.0 [1	TD_3]	- <=	>= 4		
	,		Com		Compute variable: It Cas
Target Varia	able [.]	Numeric Expr		X	Include all cases
	=	1		1	Include if case satisfies condition
Type & I	abel		💑 T=1.	0 [TD_1]	INDEX>103
			♣ T=2.	0 [TD_2] 0 [TD_3]	
GDP	*		💫 INDE	XD	
🗼 т					
♣ T=1.0 [■ T=2.0 [TD_1]	+ <	>		
					+ < > 7 8
ן U.C= ו 🐠 ו	TD_3]	- <=	>=		
INDEXD	10_3]	- <=	>=		- <= >= 4 5
INDEXD	10_3]	* _	>=		- <= >= 4 5 0
utputs:	10_3]	- <= * _	~=		- <= >= 4 5 0
utputs:	index index	- <= * _	>= ~- & T	💰 INDEXD	- <= >= 4 5 0
utputs:	INDEX 101.0	- <= ★	>= ~- T 1	NDEXD	- <= >= 4 5 0
utputs:	ID_3 INDEX 101.0 100.7	✓ GDP 28456500 18401000	>= ~= 1 1	INDEXD .00	- <= >= 4 5 0
utputs:	INDEX 101.0 100.7 103.1	✓ GDP 28456500 18401000 10854284	>= 	INDEXD .00 .00 1.00	- <= >= 4 5 0
utputs:	INDEX 101.0 100.7 103.1 99.1	✓ CDP 28456500 18401000 10854284 3863400	>= 	INDEXD .00 .00 .00 .00 .00 .00 .00 .00	
1 1 2 3 4 5	INDEX 101.0 100.7 103.1 99.1 100.3	✓ CDP 28456500 18401000 10854284 3863400 2111188	>= 	INDEXD .00 .00 .00 .00 .00 .00 .00 .00	
1 2 3 4 5 6	INDEX 101.0 100.7 103.1 99.1 100.3 101.9	✓ CDP 28456500 18401000 10854284 3863400 2111188 12364727	>= T 1 1 1 1 1 1 1 1 1	INDEXD .00 .00 .00 .00 .00 .00 .00	
1 1 2 3 4 5 6 7 7	INDEX 101.0 101.0 100.7 103.1 99.1 100.3 101.9 96.6 100.1	✓ CDP 28456500 28456500 18401000 10854284 3863400 2111188 12364727 12356400 10930125	T	INDEXD .00 .00 .00 .00 .00 .00 .00 .00 .00	
1 2 3 4 5 6 7 8	INDEX 101.0 100.7 103.1 99.1 100.3 101.9 96.6 100.1	✓ CDP 28456500 28456500 18401000 10854284 3863400 2111188 12364727 12356400 10030125 11201155	>= 	INDEXD .000 .000 .000 .000 .000 .000 .000 .000 .000 .000	
1 2 3 4 5 6 7 8 9	INDEX 101.0 100.7 103.1 99.1 100.3 101.9 96.6 100.1 102.5 105.1	CDP 28456500 28456500 18401000 10854284 3863400 2111188 12364727 12356400 10030125 11201156 40508400	T	► INDEXD .00 .00 .00 .00 .00 .00 .00 .00 .00 .00 .00 .00 .00 .00 .00 .00 .00 .00	
1 2 3 4 5 6 7 8 9 10	INDEX 101.0 101.0 100.7 103.1 99.1 100.3 101.9 96.6 100.1 102.5 105.1 100.2	CDP 28456500 28456500 18401000 10854284 3863400 2111188 12364727 12356400 10030125 11201156 49508400 11503224	T	INDEXD .00	
1 2 3 4 5 6 7 8 9 10 11 12	INDEX 101.0 100.7 103.1 99.1 100.3 101.9 96.6 100.1 102.5 105.1 100.2 104.3	CDP 28456500 28456500 18401000 10854284 3863400 2111188 12364727 12356400 10030125 11201156 49508400 11503034 15680128	T	INDEXD .000	
1 2 3 4 5 6 7 8 9 10 11 12 12	INDEX 101.0 100.7 103.1 99.1 100.3 100.3 101.9 96.6 100.1 102.5 105.1 100.2 104.3 102.0	✓ CDP 28456500 18401000 10854284 3863400 2111188 12364727 12356400 10030125 11201156 49508400 11503034 15680138 12126554	T	► INDEXD .00	

q) Estimate a logit model by using INDEXD as the dependent variable, and T and GDP as the independent variables

File I	Edit Vie	ew	Data	Transform	n Analyze	Graphs	Utilities	Extensions	Window	Help		
					Power	Analysis		•	IBM SPSS	Statistics	Data Editor	
		1			Report Descrij Bayesi Tables	s otive Statis an Statistic	stics cs	•	Q			
DEX	🧳 G	DP		뤚 Т	Compa	ire Means		•	NDEXD	var	var	var
96.6	123	5640	0	1	Genera	General Linear Model Generalized Linear Models						
100.1	100	3012	5	1	Genera							
102.5	112	0115	6	1	Mixed	Models			.00			
105.1	495	0840	0	1	Regres	sion			Aut	omatic I	inear Model	ing
100.2	115	0303	4	1	Logline	ear		•	Lin	ear		
104.3	156	8013	8	1	Neural	Networks		•	Cur	ve Estim	ation	
109.0	131	2685	4	1	Classif	у			Par	tial Least	Squares	
100.4	36	3441	2	1	Dimen	sion Reduc	ction		R Bin	ary Logis	tic	
102.8	107	4778	٩	1	Scale							

	1	Logistic Regressi	on	
 ✓ INDEX ✓ GDP ✓ T 	Depen Depen Block 1 of 1 Previous	dent: DEXD	Next	Categorical Save Options
-	► >a*b>	Block 1 of 1 GDP T		Style Bootstrap
-	Method:	Enter	\$	
	Select	ion Variable:	Rule	
? F	Reset Paste		Can	ocel OK

Outputs:

Classification Table^a

			Predicted			
			IND	EXD	Percentage	
Observed		.00	1.00	Correct		
Step 1	INDEXD	.00	57	8	87.7	
		1.00	22	13	37.1	
	Overall Pe	ercentage			70.0	

a. The cut value is .500

Variables in the Equation

		В	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	GDP	.000	.000	1.721	1	.190	1.000
	Т	.704	.283	6.189	1	.013	2.021
	Constant	-2.404	.661	13.223	1	.000	.090

r) Predict the value of INDEXD when T = 1, 2, and 3 and GDP = 30,000000

Step 1: Input the value of T and GDP into the datafile.

	🧳 INDEX	🖋 GDP	🔒 т	🔒 INDEXD	
07	110.9	22505600	2	1.00	
97	110.9	22303000		1.00	
98	105.9	18708046	3	1.00	
99	101.5	3809191	3	.00	
100	100.3	8120121	3	.00	
101	· · · ·	3000000	1	· · · ·	
102		3000000	2		
103		3000000	3		

Step 2: Re-estimate the model, and request the unstandardised predicted value of INDEXD.

Outputs:

	🛷 INDEX	🧳 GDP	🕹 Т	💦 INDEXD	PRE_1	💑 PGR_1
96	104.3	2288595	3	1.00	.44239	.00
97	110.9	22505600	3	1.00	.57760	1.00
98	105.9	18708046	3	1.00	.55247	1.00
99	101.5	3809191	3	.00	.45251	.00
100	100.3	8120121	3	.00	.48140	.00
101		3000000	1		.29063	.00
102		3000000	2	•	.45294	.00
103		3000000	3	•	.62591	1.00
104						