



SDSC 2102

STATISTICAL METHODS AND DATA ANALYSIS

Lab 3,4 – Linear Regression Analysis using SPSS



Aims

- ❑ A. Learn the process of model training and model testing
 - ❑ Split dataset to training set and test set;
 - ❑ Make prediction for the test set.

- ❑ B. Learn to perform simple & multiple linear regression
 - ❑ Check the correlations of the variables;
 - ❑ Compute regression coefficients;
 - ❑ Understand the role of R square;
 - ❑ Able to recognize multicollinearity.

Model Description

- Given a dataset $\{(x_i, y_i), i = 1, \dots, n\}$ and we want to find the relationship between x and y .
- In linear regression model, a line is used to fit the relationship.

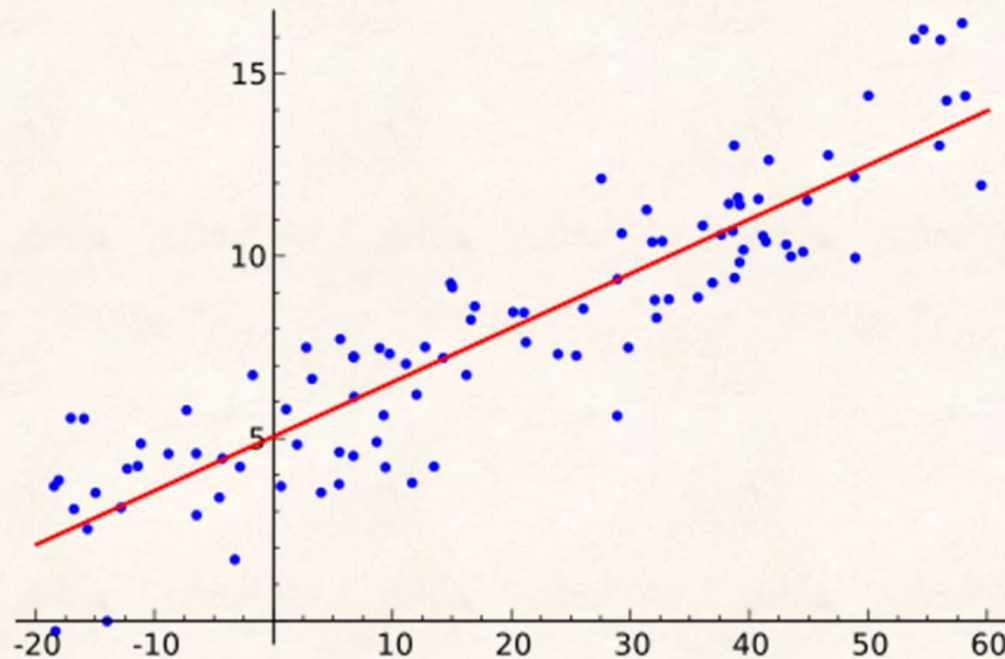


Figure source: Wikipedia: Simple linear regression

Model Description

- Simple (univariate) Linear Regression:

$$\hat{y} = \theta_0 + \theta_1 x$$

$\hat{y}$: predicted value of the response

θ_0 : y -intercept of the regression line

θ_1 : slope of the regression line

x : predictor, explanatory or independent variable

Model Description

□ Simple Linear Regression:

$$\begin{aligned} y_i &= \hat{y} + e_i \\ &= \theta_0 + \theta_1 x_i + e_i \end{aligned}$$

y_i : true value of the response

e_i : error term (residual) satisfying $\sum_{i=1}^n e_i = 0$

Model Description

□ Coefficient of Determination: R Square

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{SSE}{SST}$$

SSE: the sum of squared error

SST: the sum of squared total

$R^2 = 0$: Response cannot be predicted from the independent variable.

$R^2 = 1$: Response can be perfectly predicted without error from the independent variable.

Training Set and Test Set

- **Training dataset** is the set, on which the model is initially fitted and the model parameters are found.
- **Test dataset** is to judge the performance of the selected model.
- We can use 80% data as training set and the left 20% data are used as test set.

Model Description

- Multiple Linear Regression:

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$$

- Adjusted R Square:

$$R_a^2 = R^2 - \frac{k(1 - R^2)}{n - k - 1}$$

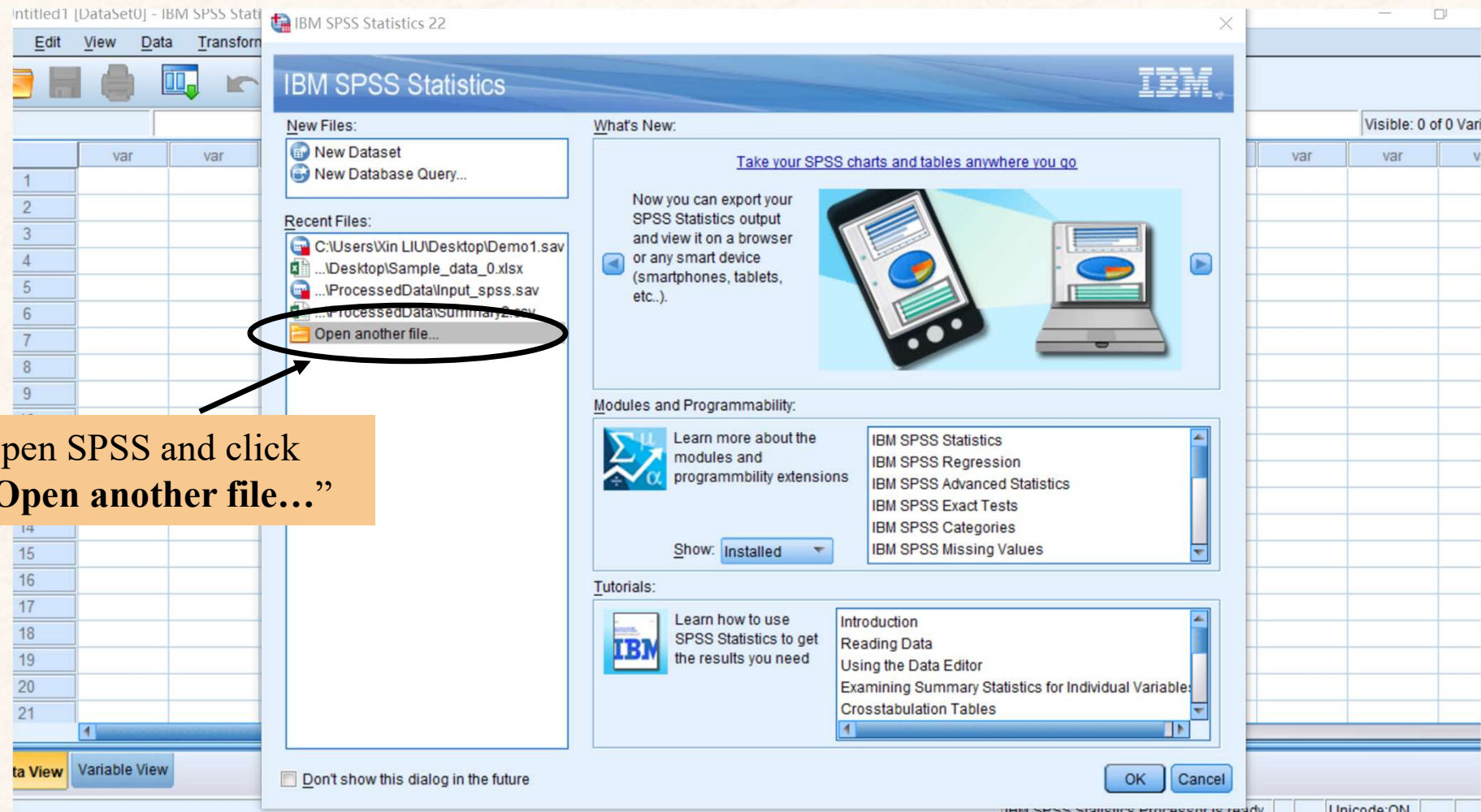
- Variance Inflation Factor (VIF):

A index to check multicollinearity.

If $VIF \geq 10$, model has severe multicollinearity, which may increase the variance of the coefficient estimates.

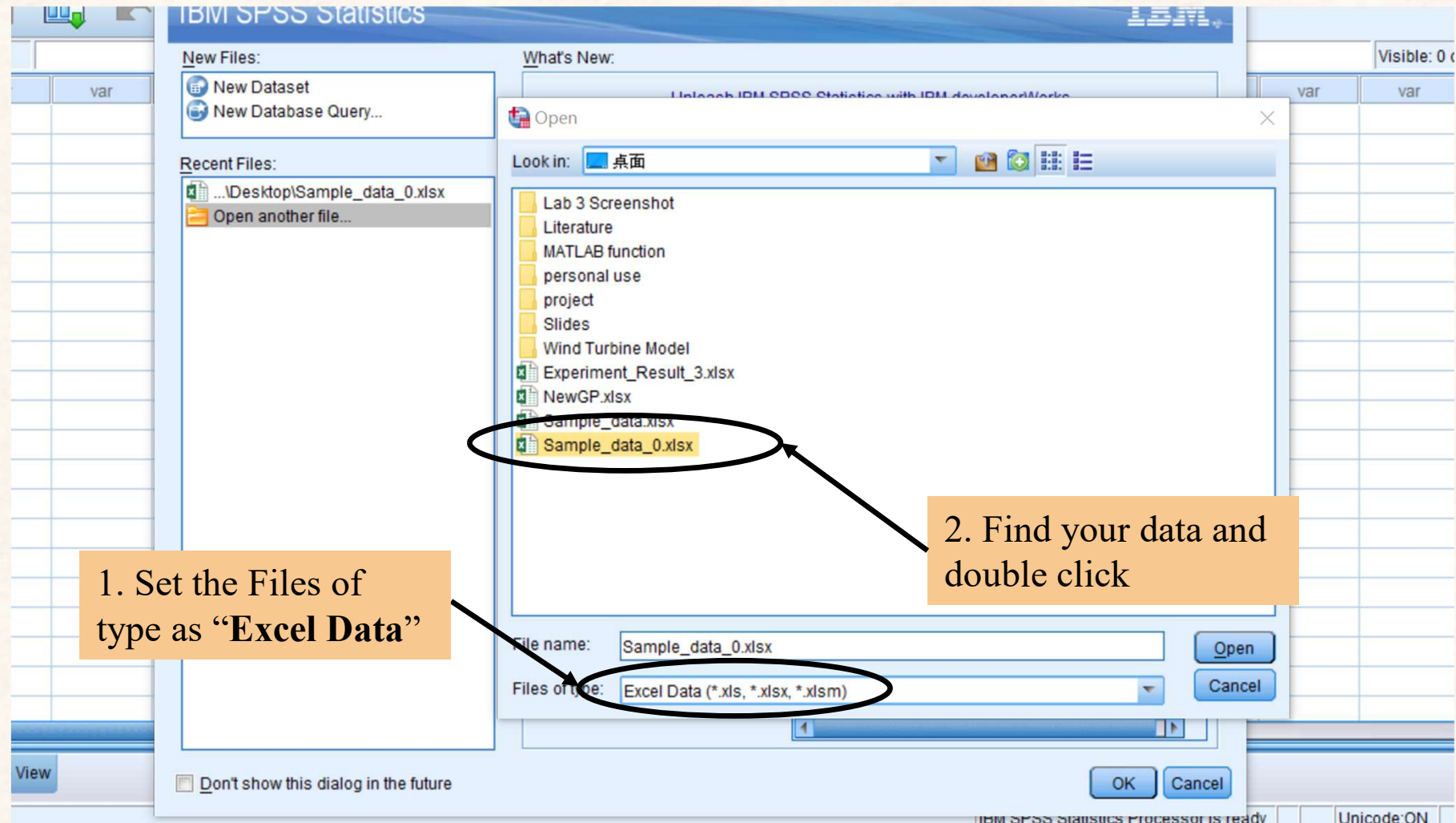
Part A. Split dataset

□ Import Data



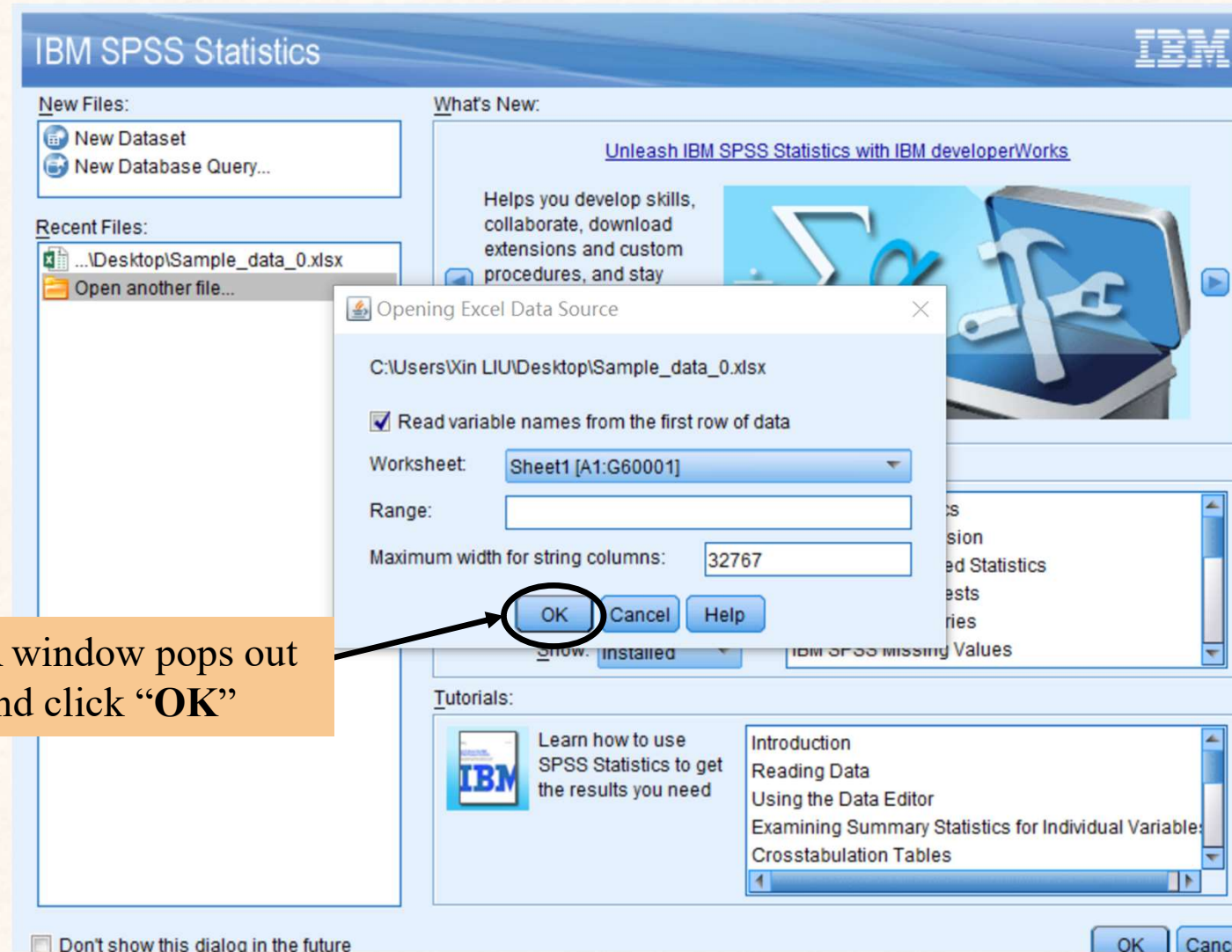
Part A. Split dataset

□ Import Data



Part A. Split dataset

□ Import Data



Part A. Split dataset

- Now the data is imported.

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Visible: 7 of 7 Variables

	Time	Response	Response100000	Input1	Input2	Input3	Input4	var
1	01-Oct-2008	.0003535419957635	35.3541995763512300	5.7272653579711910	12.2393522262573240	-.4781374831994374	-29.3236083984375000	
2	01-Oct-2008	.0003817622656483	38.1762265648344000	5.7358798980712890	12.2353401184082030	-.4781445364157359	-31.3377075195312500	
3	01-Oct-2008	.0004901928987136	49.0192898713623900	5.2923078536987305	12.2313280105590820	-.4781515896320343	-31.1883449554443360	
4	01-Oct-2008	.0004253186862309	42.5318686230933000	5.5644450187683105	12.2273168563842770	-.4781586428483327	-22.2278347015380860	
5	01-Oct-2008	.0003716644736330	37.1664473633003600	5.8365821838378910	12.2233047485351560	-.4781656960646312	-25.4763774871826170	
6	01-Oct-2008	.0003280453887398	32.8045388739753040	6.1087193489074710	12.2192926406860350	-.4781727691491445	-35.8977241516113300	
7	01-Oct-2008	.0003090853467528	30.9085346752793600	6.3077316284179690	12.2152805328369140	-.4781798223654429	-26.1010131835937500	
8	01-Oct-2008	.0003585741727826	35.8574172782604000	6.0742130279541020	12.2112693786621100	-.4781868755817413	-14.6254701614379880	
9	01-Oct-2008	.0003358370583409	33.5837058340910700	6.2793655395507810	12.2123937606811520	-.4781939287980397	-25.8846435546875000	
10	01-Oct-2008	.0003907306916921	39.0730691692115000	6.0364079475402830	12.2533969879150390	-.4782009820143382	-22.4881954193115230	
11	01-Oct-2008	.0004415300483621	44.1530048362149260	5.8574786186218260	12.2944002151489260	-.4782080352306366	-3.8344509601593018	
12	01-Oct-2008	.0004039063300824	40.3906330082370960	6.0850696563720700	12.3354034423828120	-.4782150884469350	5.9199233055114750	
13	01-Oct-2008	.0003848553583987	38.4855358398694700	6.1554503440856930	12.3764066696167000	-.4782221416632335	16.8376522064209000	
14	01-Oct-2008	.0003667757911329	36.6775791132917000	6.2258310317993160	12.4174108505249020	-.4782291948795319	1.2255277633666992	
15	01-Oct-2008	.0003441899034093	34.4189903409301200	6.3429355621337890	12.4584140777587890	-.4782362580299377	-3.7134051322937010	
16	01-Oct-2008	.0002949460057972	29.4946005797247040	6.7477374076843260	12.4994173049926760	-.4782433112462362	-1.2015984058380127	
17	01-Oct-2008	.0003022158733150	30.2215873315000180	6.5956001281738280	12.5404205322265620	-.4782503743966420	-.2330269664525986	
18	01-Oct-2008	.0003285147653682	32.8514765368215400	6.4434628486633300	12.5814237594604500	-.4782574276129405	-8.3876247406005860	
19	01-Oct-2008	.0003629734263869	36.2973426386911840	6.2603383064270020	12.6224269866943360	-.4782644907633464	-4.0000000000000000	
20	01-Oct-2008	.0004120962422249	41.2096242224911300	6.0272130966186520	12.6634302139282230	-.4782715439796448	-4.0000000000000000	
21	01-Oct-2008	.0002794256691670	27.9425669167011180	6.8901643753051760	12.7044343948364260	-.4782785971959432	-4.4861187934875490	

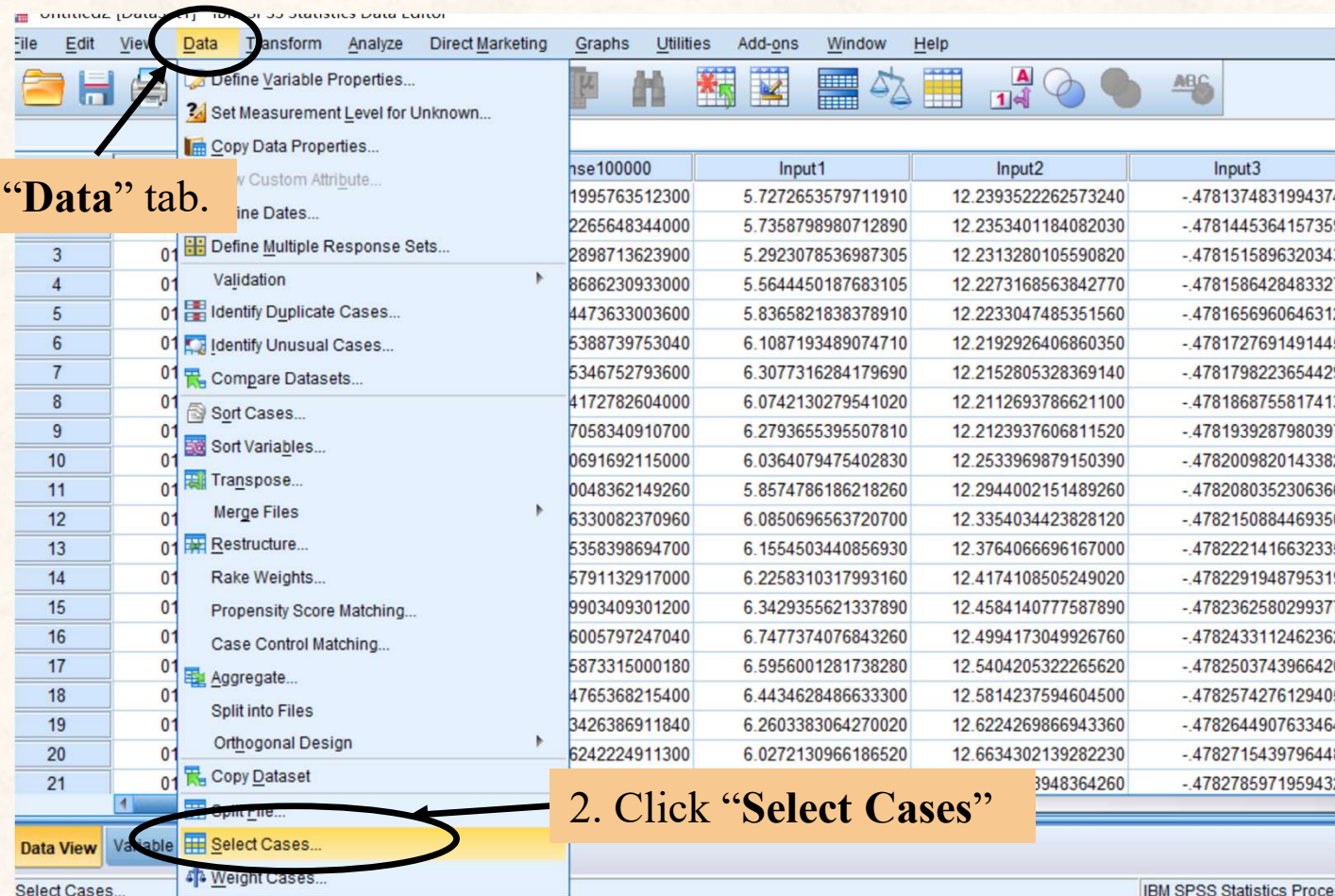
Data View Variable View

IBM SPSS Statistics Processor is ready Unicode: ON

Part A. Split dataset

- Split Dataset: Label samples to indicate in which set the sample is.

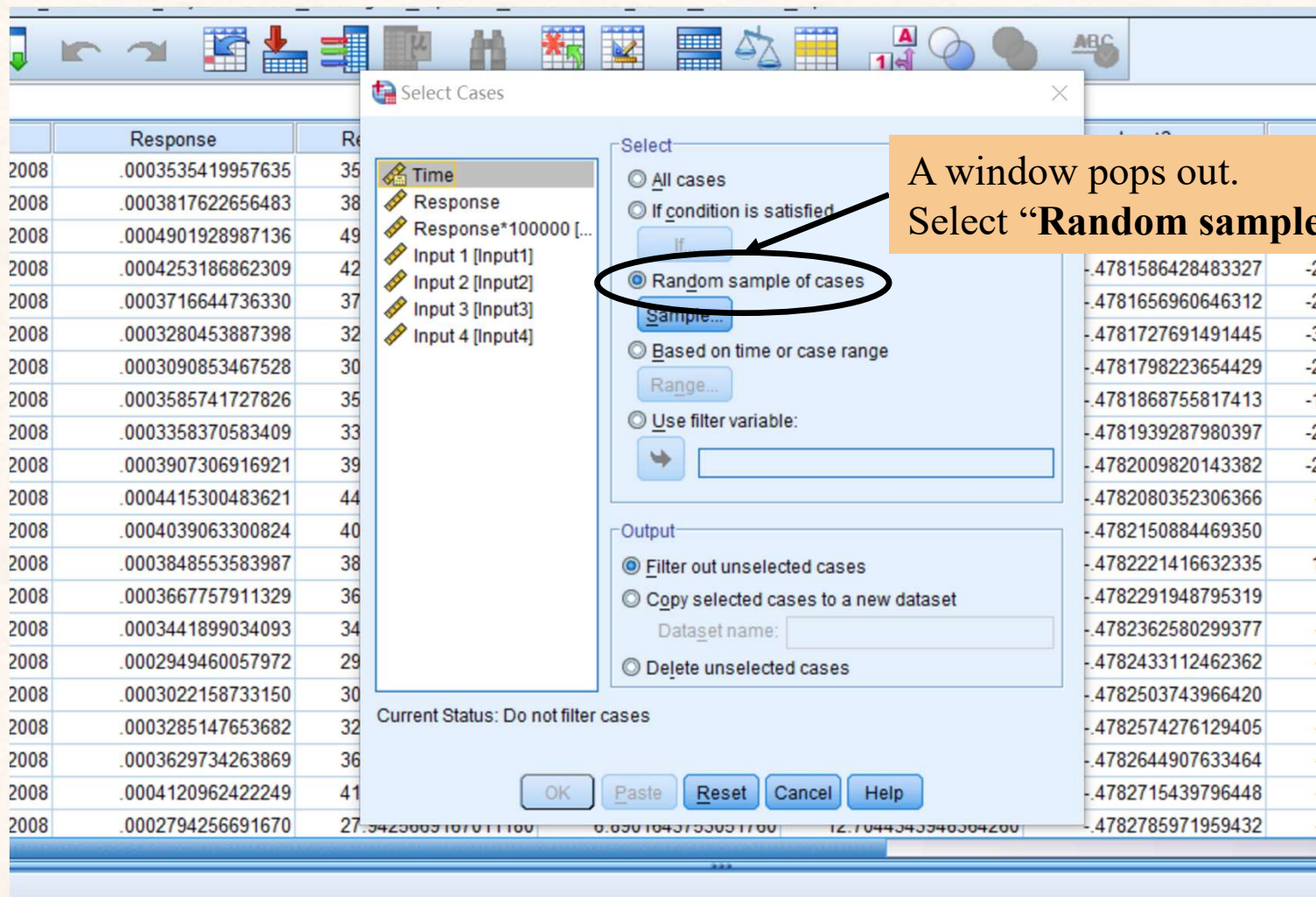
1. Click “Data” tab.



2. Click “Select Cases”

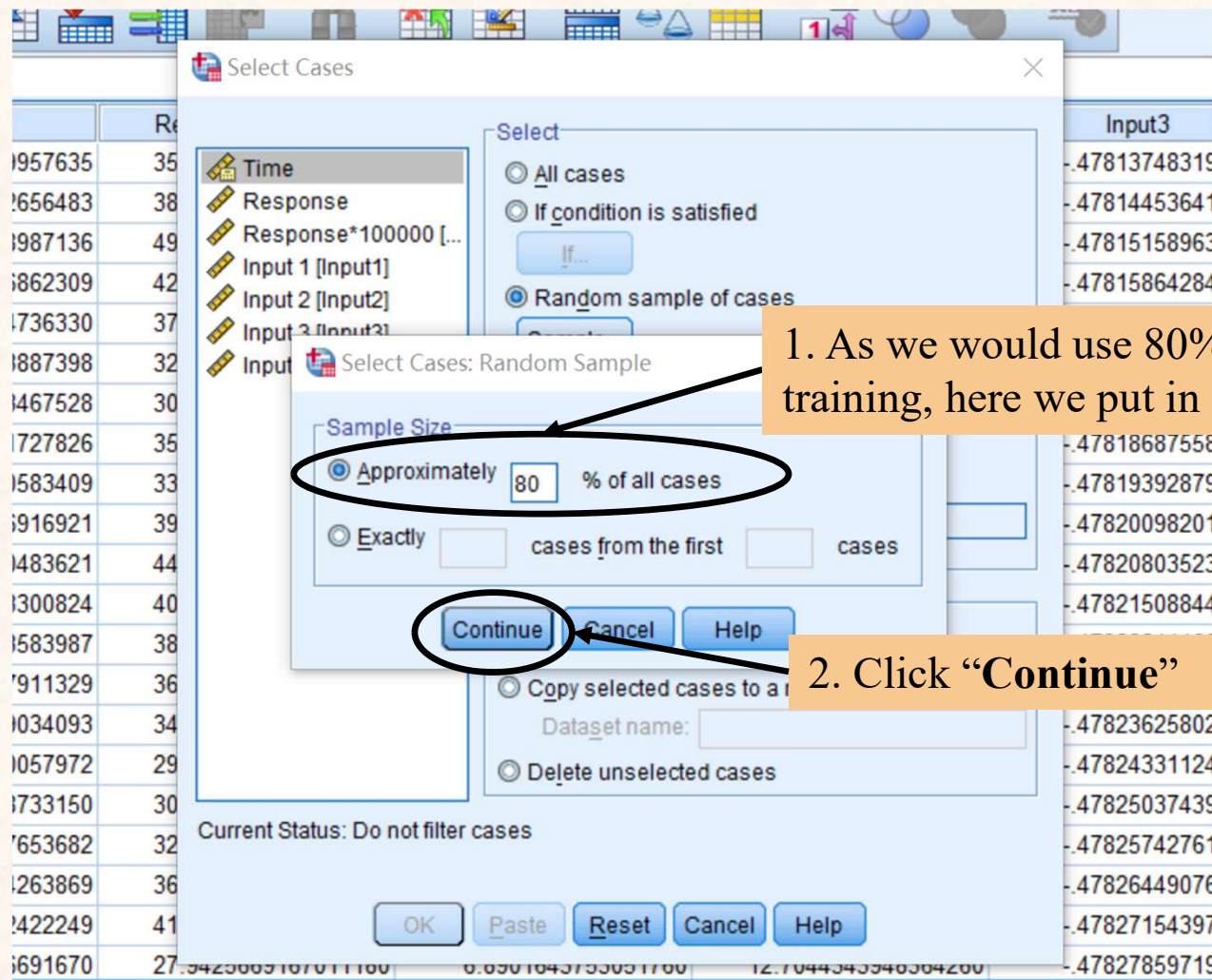
Part A. Split dataset

□ Split Dataset: Label samples



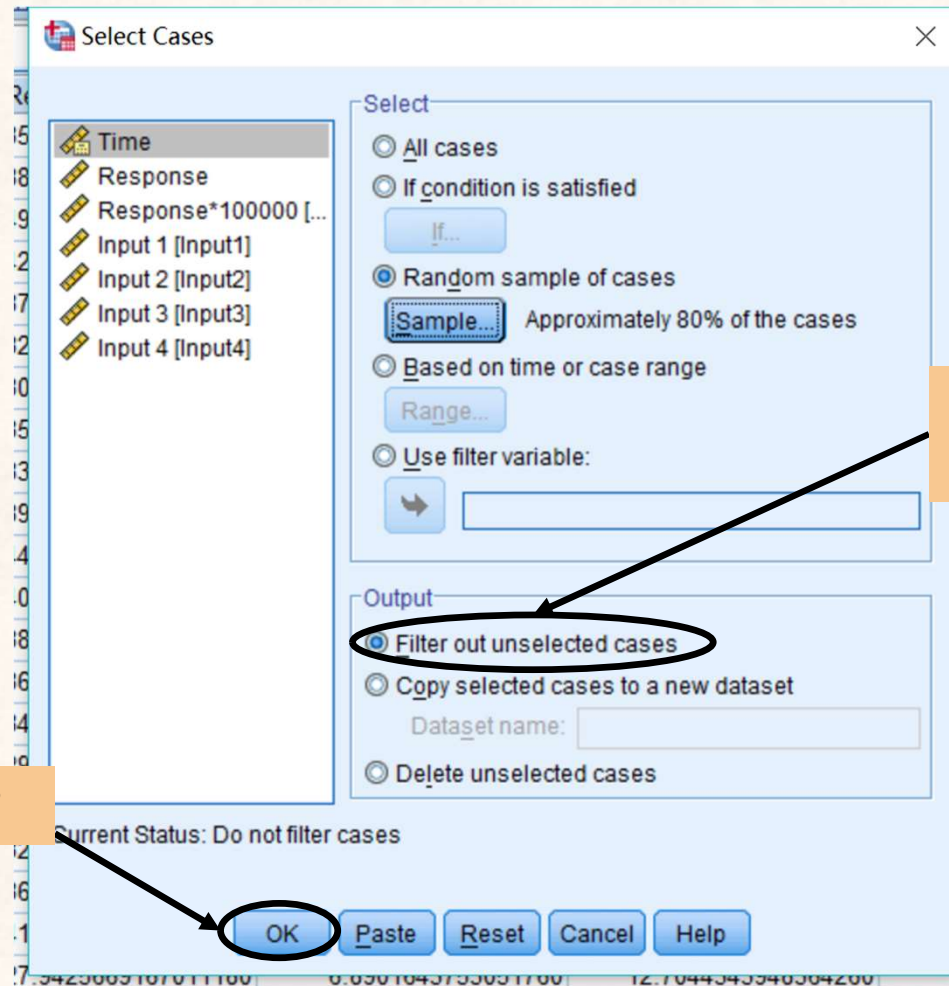
Part A. Split dataset

□ Split Dataset: Label samples



Part A. Split dataset

□ Split Dataset: Label samples



1. Select “Filter out unselected cases”

2. Click “OK”

Part A. Split dataset

□ Split Dataset: Label samples

At the far right, a new variable named “filter_\$” is generated, where 1 denotes the training set samples and 0 denotes the test set samples.

The test set samples are stroked out.

	Time	Response	Response100000	Input1	Input2	Input3	Input4	filter_\$
1	01-Oct-2008	.0003535419957635	35.3541995763512300	5.7272653579711910	12.2393522262573240	-4781374831994374	-29.32360839843750	1
2	01-Oct-2008	.0003817622656483	38.1762265648344000	5.7358798980712890	12.2352101181082020	4781115261157250	21.3277075195312500	1
3	01-Oct-2008	.0004901928987136	49.0192898713623900	5.2923078536987305			955.443360	1
4	01-Oct-2008	.0004253186862309	42.5318686230933000	5.5644450187683105			7015380860	1
5	01-Oct-2008	.0003716644736330	37.1664473633003600	5.8365821838378910			4871826170	1
6	01-Oct-2008	.0003280453887398	32.8045388739753040	6.1087193489074710			1516113300	1
7	01-Oct-2008	.0003090853467528	30.9085346752793600	6.3077316284179690			1835937500	1
8	01-Oct-2008	.0003585741727826	35.8574172782604000	6.0742130279541020			1614379880	1
9	01-Oct-2008	.0003358370583409	33.5837058340910700	6.2793655395507810			5546875000	1
10	01-Oct-2008	.0003907306916921	39.0730691692115000	6.0364079475402830			4193115230	1
11	01-Oct-2008	.0004415300483621	44.1530048362149260	5.8574786186218260			9601593018	1
12	01-Oct-2008	.0004039063300824	40.3906330082370960	6.0850696563720700			3055114750	1
13	01-Oct-2008	.0003848553583987	38.4855358398694700	6.1554503440856930			2064209000	1
14	01-Oct-2008	.0003667757911329	36.6775791132917000	6.2258310317993160			7633666992	1
15	01-Oct-2008	.0003441899034093	34.4189903409301200	6.3429355621337890	12.4584140777587890	-4782362580299377	-3.7134051322937010	1
16	01-Oct-2008	.0002949460057972	29.4946005797247040	6.7477374076843260	12.4994173049926760	-4782433112462362	-1.2015984058380127	0
17	01-Oct-2008	.0003022158733150	30.2215873315000180	6.5956001281738280	12.5404205322265620	-4782503743966420	-2330269664525986	1
18	01-Oct-2008	.0003285147653682	32.8514765368215400	6.4434628486633300	12.5814237594604500	-4782574276129405	-8.3876247406005860	1
19	01-Oct-2008	.0003629734263869	36.2973426386911840	6.2603383064270020	12.6224269866943360	-4782644907633464	-4.0000000000000000	1
20	01-Oct-2008			0272130966186520	12.6634302139282230	-4782715439796448	-4.0000000000000000	1
21	01-Oct-2008			8901643753051760	12.7044343948364260	-4782785971959432	-4.4861187934875490	1

Part A. Split dataset

- Split out the test set according to the label.

1. Click “Data” tab.

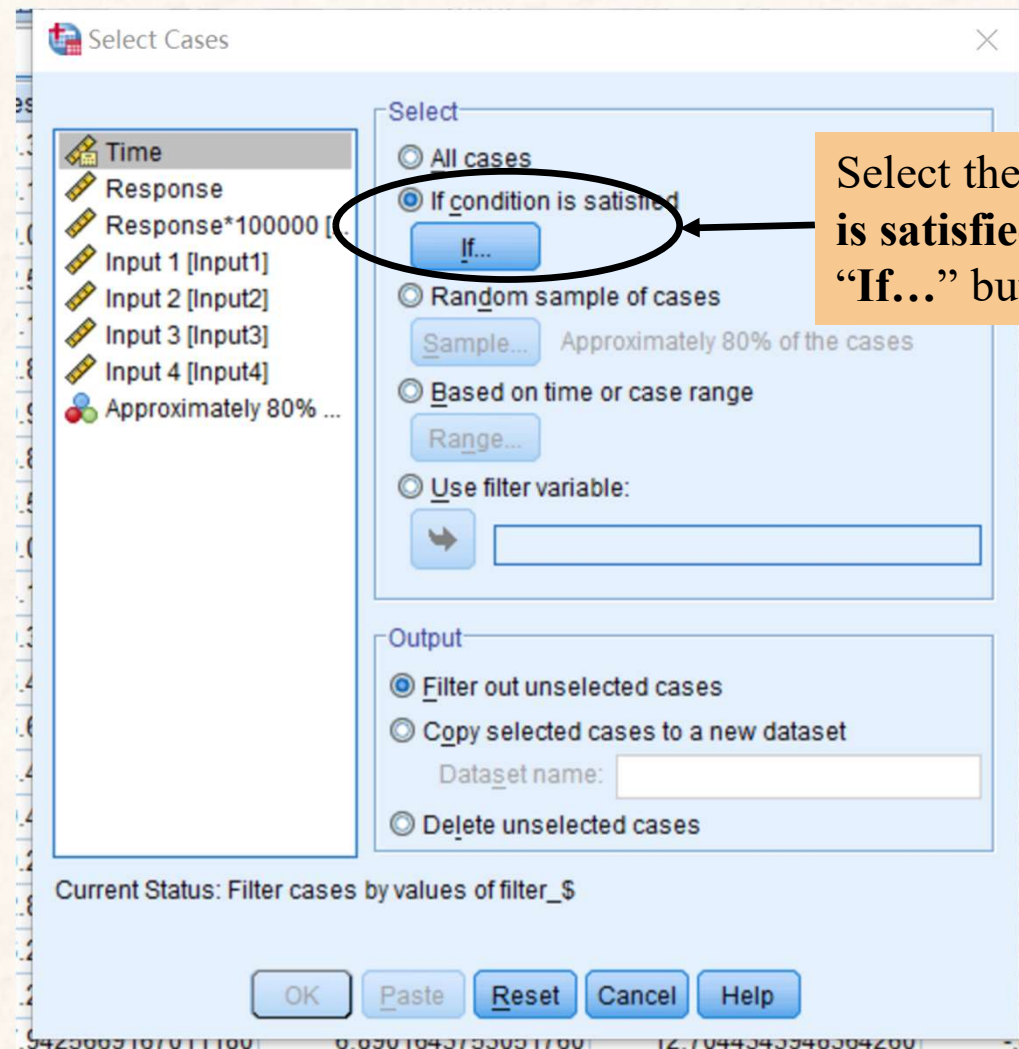
The screenshot shows the IBM SPSS Data Editor interface. The 'Data' menu is open, and the 'Select Cases...' option is highlighted. An arrow points from the 'Data' menu to the 'Select Cases...' option. Another arrow points from the 'Select Cases...' option to the text '2. Click “Select Cases”'. The background shows a data table with columns 'Input1' and 'Input2'.

	Input1	Input2
995763512300	5.7272653579711910	12.2393522262573240
265648344000	5.7358798980712890	12.2353401184082030
398713623900	5.2923078536987305	12.2313280105590820
586230933000	5.5644450187683105	12.2273168563842770
473633003600	5.8365821838378910	12.2233047485351560
388739753040	6.1087193489074710	12.2192926406860350
346752793600	6.3077316284179690	12.2152805328369140
172782604000	6.0742130279541020	12.2112693786621100
058340910700	6.2793655395507810	12.2123937606811520
591692115000	6.0364079475402830	12.2533969879150390
048362149260	5.8574786186218260	12.2944002151489260
330082370960	6.0850696563720700	12.3354034423828120
358398694700	6.1554503440856930	12.3764066696167000
791132917000	6.2258310317993160	12.4174108505249020
903409301200	6.3429355621337890	12.4584140777587890
005797247040	6.7477374076843260	12.4994173049926760
873315000180	6.5956001281738280	12.5404205322265620
765368215400	6.4434628486633300	12.5814237594604500
426386911840	6.2603383064270020	12.6224269866943360
242224911300	6.0272130966186520	12.6634302139282230
		12.7044343948364260

2. Click “Select Cases”

Part A. Split dataset

- Split out the test set according to the label.



Part A. Split dataset

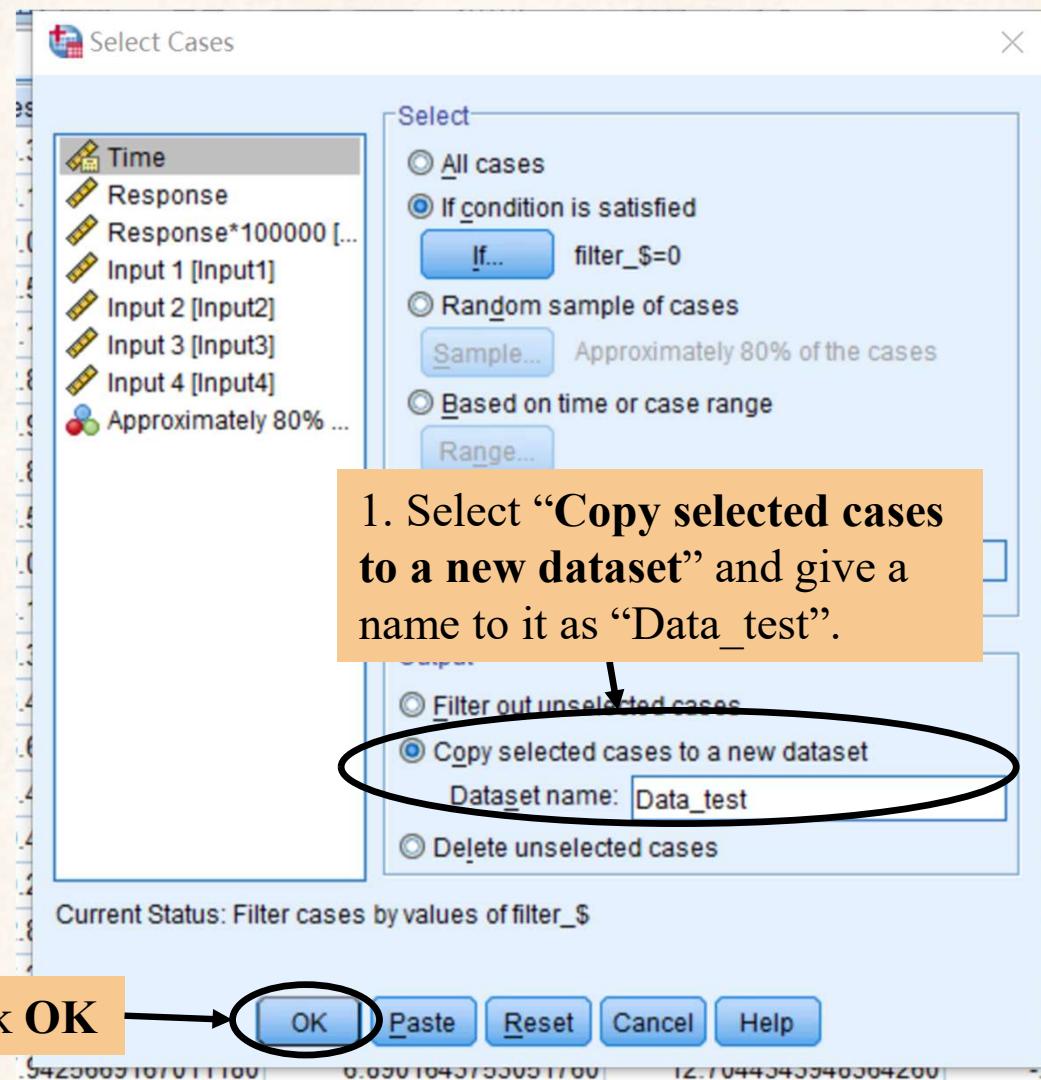
The image shows two sequential screenshots of the SPSS 'Select Cases: If' dialog box, illustrating the steps to split a dataset.

Step 1: In the first screenshot, the variable 'Approximately 80% ...' is selected in the list on the left. An arrow points from this selection to the 'filter_\$' textbox in the center. A text box explains: 'A new window pops out. 1. Double click the new created variable at the left side, then "filter_\$" will be shown in this textbox above.'

Step 2: In the second screenshot, the variable 'filter_\$=0' is entered in the 'filter_\$' textbox. An arrow points from the text box '2. Then type " =0 "' to the 'filter_\$' textbox. The dialog box also shows a calculator keypad and a list of function groups.

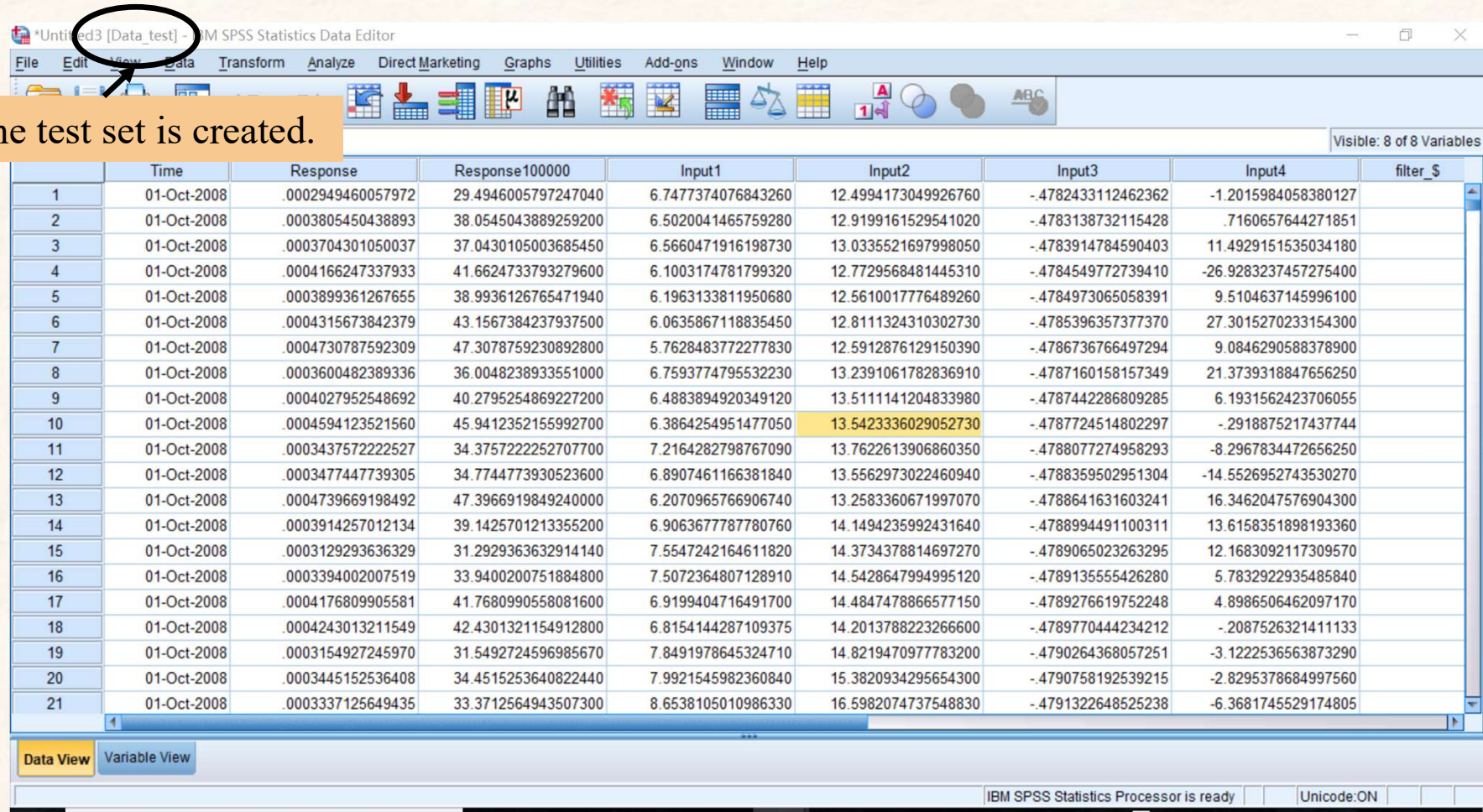
Step 3: In the third screenshot, the 'Continue' button is highlighted. An arrow points from the text box '3. Click "Continue"' to the 'Continue' button.

Part A. Split dataset



Part A. Split dataset

Now the test set is created.



Visible: 8 of 8 Variables

	Time	Response	Response100000	Input1	Input2	Input3	Input4	filter_\$
1	01-Oct-2008	.0002949460057972	29.4946005797247040	6.7477374076843260	12.4994173049926760	-.4782433112462362	-1.2015984058380127	
2	01-Oct-2008	.0003805450438893	38.0545043889259200	6.5020041465759280	12.9199161529541020	-.4783138732115428	.7160657644271851	
3	01-Oct-2008	.0003704301050037	37.0430105003685450	6.5660471916198730	13.0335521697998050	-.4783914784590403	11.4929151535034180	
4	01-Oct-2008	.0004166247337933	41.6624733793279600	6.1003174781799320	12.7729568481445310	-.4784549772739410	-26.9283237457275400	
5	01-Oct-2008	.0003899361267655	38.9936126765471940	6.1963133811950680	12.5610017776489260	-.4784973065058391	9.5104637145996100	
6	01-Oct-2008	.0004315673842379	43.1567384237937500	6.0635867118835450	12.8111324310302730	-.4785396357377370	27.3015270233154300	
7	01-Oct-2008	.0004730787592309	47.3078759230892800	5.7628483772277830	12.5912876129150390	-.4786736766497294	9.0846290588378900	
8	01-Oct-2008	.0003600482389336	36.0048238933551000	6.7593774795532230	13.2391061782836910	-.4787160158157349	21.3739318847656250	
9	01-Oct-2008	.0004027952548692	40.2795254869227200	6.4883894920349120	13.5111141204833980	-.4787442286809285	6.1931562423706055	
10	01-Oct-2008	.0004594123521560	45.9412352155992700	6.3864254951477050	13.5423336029052730	-.4787724514802297	-.2918875217437744	
11	01-Oct-2008	.0003437572222527	34.3757222252707700	7.2164282798767090	13.7622613906860350	-.4788077274958293	-8.2967834472656250	
12	01-Oct-2008	.0003477447739305	34.7744773930523600	6.8907461166381840	13.5562973022460940	-.4788359502951304	-14.5526952743530270	
13	01-Oct-2008	.0004739669198492	47.3966919849240000	6.2070965766906740	13.2583360671997070	-.4788641631603241	16.3462047576904300	
14	01-Oct-2008	.0003914257012134	39.1425701213355200	6.9063677787780760	14.1494235992431640	-.4788994491100311	13.6158351898193360	
15	01-Oct-2008	.0003129293636329	31.2929363632914140	7.5547242164611820	14.3734378814697270	-.4789065023263295	12.1683092117309570	
16	01-Oct-2008	.0003394002007519	33.9400200751884800	7.5072364807128910	14.5428647994995120	-.4789135555426280	5.7832922935485840	
17	01-Oct-2008	.0004176809905581	41.7680990558081600	6.9199404716491700	14.4847478866577150	-.4789276619752248	4.8986506462097170	
18	01-Oct-2008	.0004243013211549	42.4301321154912800	6.8154144287109375	14.2013788223266600	-.4789770444234212	-.2087526321411133	
19	01-Oct-2008	.0003154927245970	31.5492724596985670	7.8491978645324710	14.8219470977783200	-.4790264368057251	-3.1222536563873290	
20	01-Oct-2008	.0003445152536408	34.4515253640822440	7.9921545982360840	15.3820934295654300	-.4790758192539215	-2.8295378684997560	
21	01-Oct-2008	.0003337125649435	33.3712564943507300	8.6538105010986330	16.5982074737548830	-.4791322648525238	-6.3681745529174805	

Data View Variable View

IBM SPSS Statistics Processor is ready Unicode:ON

Part A. Split dataset

1. Move back to the “Dataset1”

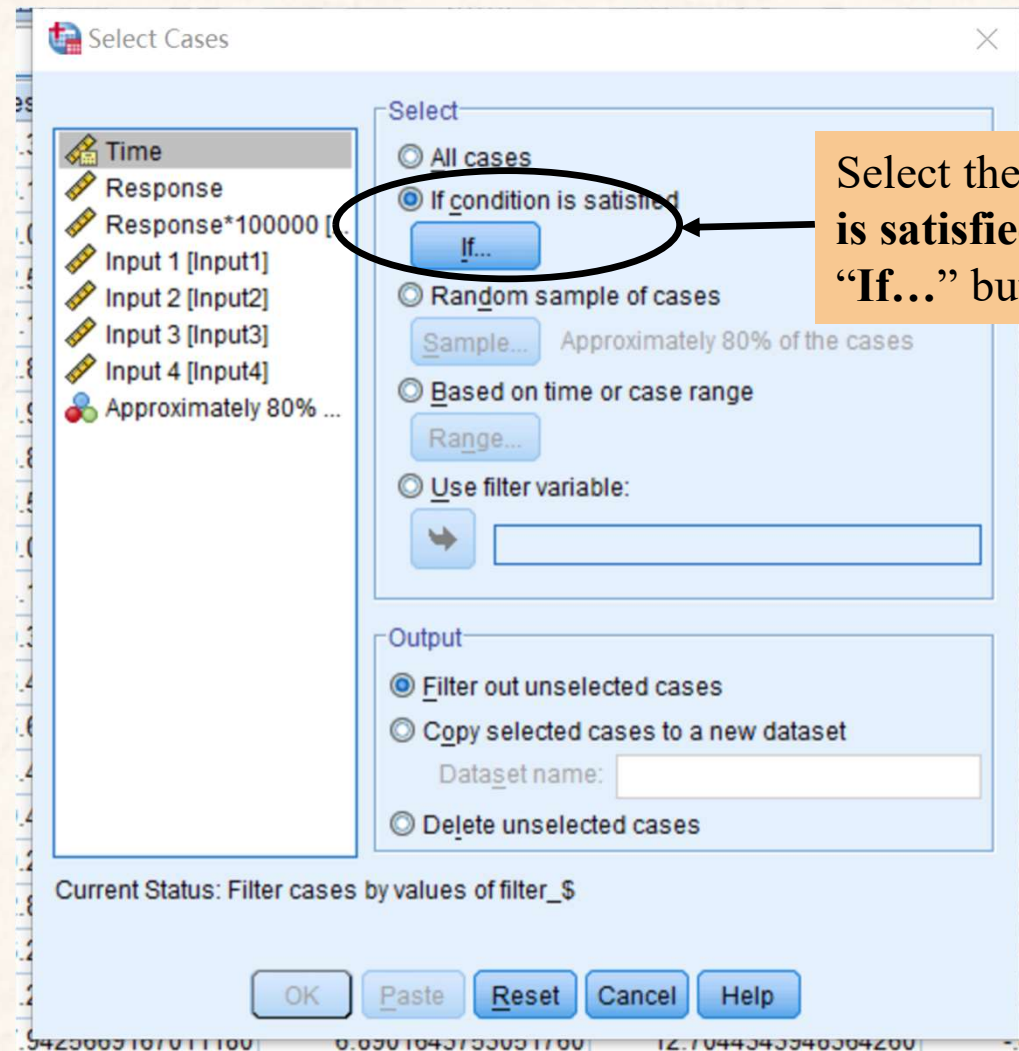
2. Click “Data” tab.

The screenshot shows the IBM SPSS Data Editor interface. The 'Data' menu is open, and the 'Select Cases...' option is highlighted. The background data table is visible, showing columns for 'Case', 'Input1', and 'Input2'.

Case	Input1	Input2
100000		
995763512300	5.7272653579711910	12.2393522262573240
265648344000	5.7358798980712890	12.2353401184082030
398713623900	5.2923078536987305	12.2313280105590820
586230933000	5.5644450187683105	12.2273168563842770
473633003600	5.8365821838378910	12.2233047485351560
388739753040	6.1087193489074710	12.2192926406860350
346752793600	6.3077316284179690	12.2152805328369140
172782604000	6.0742130279541020	12.2112693786621100
058340910700	6.2793655395507810	12.2123937606811520
591692115000	6.0364079475402830	12.2533969879150390
048362149260	5.8574786186218260	12.2944002151489260
330082370960	6.0850696563720700	12.3354034423828120
358398694700	6.1554503440856930	12.3764066696167000
791132917000	6.2258310317993160	12.4174108505249020
903409301200	6.3429355621337890	12.4584140777587890
005797247040	6.7477374076843260	12.4994173049926760
873315000180	6.5956001281738280	12.5404205322265620
765368215400	6.4434628486633300	12.5814237594604500
426386911840	6.2603383064270020	12.6224269866943360
242224911300	6.0272130966186520	12.6634302139282230
		12.7044343948364260

3. Click “Select Cases”

Part A. Split dataset



Part A. Split dataset

The image shows two sequential screenshots of the SPSS 'Select Cases: If' dialog box, illustrating the steps to split a dataset.

Step 1: In the first screenshot, the variable 'Approximately 80% ...' is selected in the list on the left. An arrow points from the text '1. Double click the new created variable at the left side as before.' to this variable.

Step 2: In the second screenshot, the variable 'filter_\$' is entered in the 'filter_\$' text box. An arrow points from the text '2. Then type “=1”.' to the text box, which now contains 'filter_\$=1'.

Step 3: In the third screenshot, the 'Continue' button is highlighted. An arrow points from the text '3. Click “Continue”' to this button.

Part A. Split dataset

Now the training set is created.

The screenshot shows the IBM SPSS Statistics Data Editor window. The menu bar includes File, Edit, View, Data, Transform, Analyze, Direct Marketing, Graphs, Utilities, Add-ons, Window, and Help. The toolbar contains various icons for file operations, data manipulation, and analysis. The data grid displays 21 rows of data with 8 visible variables: Time, Response, Response100000, Input1, Input2, Input3, Input4, and filter_\$. The status bar at the bottom indicates 'IBM SPSS Statistics Processor is ready' and 'Unicode:ON'.

	Time	Response	Response100000	Input1	Input2	Input3	Input4	filter_ \$
1	01-Oct-2008	.0003535419957635	35.3541995763512300	5.7272653579711910	12.2393522262573240	-.4781374831994374	-29.3236083984375000	
2	01-Oct-2008	.0003817622656483	38.1762265648344000	5.7358798980712890	12.2353401184082030	-.4781445364157359	-31.3377075195312500	
3	01-Oct-2008	.0004901928987136	49.0192898713623900	5.2923078536987305	12.2313280105590820	-.4781515896320343	-31.1883449554443360	
4	01-Oct-2008	.0004253186862309	42.5318686230933000	5.5644450187683105	12.2273168563842770	-.4781586428483327	-22.2278347015380860	
5	01-Oct-2008	.0003716644736330	37.1664473633003600	5.8365821838378910	12.2233047485351560	-.4781656960646312	-25.4763774871826170	
6	01-Oct-2008	.0003280453887398	32.8045388739753040	6.1087193489074710	12.2192926406860350	-.4781727691491445	-35.8977241516113300	
7	01-Oct-2008	.0003090853467528	30.9085346752793600	6.3077316284179690	12.2152805328369140	-.4781798223654429	-26.1010131835937500	
8	01-Oct-2008	.0003585741727826	35.8574172782604000	6.0742130279541020	12.2112693786621100	-.4781868755817413	-14.6254701614379880	
9	01-Oct-2008	.0003358370583409	33.5837058340910700	6.2793655395507810	12.2123937606811520	-.4781939287980397	-25.8846435546875000	
10	01-Oct-2008	.0003907306916921	39.0730691692115000	6.0364079475402830	12.2533969879150390	-.4782009820143382	-22.4881954193115230	
11	01-Oct-2008	.0004415300483621	44.1530048362149260	5.8574786186218260	12.2944002151489260	-.4782080352306366	-3.8344509601593018	
12	01-Oct-2008	.0004039063300824	40.3906330082370960	6.0850696563720700	12.3354034423828120	-.4782150884469350	5.9199233055114750	
13	01-Oct-2008	.0003848553583987	38.4855358398694700	6.1554503440856930	12.3764066696167000	-.4782221416632335	16.8376522064209000	
14	01-Oct-2008	.0003667757911329	36.6775791132917000	6.2258310317993160	12.4174108505249020	-.4782291948795319	1.2255277633666992	
15	01-Oct-2008	.0003441899034093	34.4189903409301200	6.3429355621337890	12.4584140777587890	-.4782362580299377	-3.7134051322937010	
16	01-Oct-2008	.0003022158733150	30.2215873315000180	6.5956001281738280	12.5404205322265620	-.4782503743966420	-.2330269664525986	
17	01-Oct-2008	.0003285147653682	32.8514765368215400	6.4434628486633300	12.5814237594604500	-.4782574276129405	-8.3876247406005860	
18	01-Oct-2008	.0003629734263869	36.2973426386911840	6.2603383064270020	12.6224269866943360	-.4782644907633464	-4.0000000000000000	
19	01-Oct-2008	.0004120962422249	41.2096242224911300	6.0272130966186520	12.6634302139282230	-.4782715439796448	-4.0000000000000000	
20	01-Oct-2008	.0002794256691670	27.9425669167011180	6.8901643753051760	12.7044343948364260	-.4782785971959432	-4.4861187934875490	
21	01-Oct-2008	.0003922854038970	39.2285403897000650	6.1874928474426270	12.7454376220703120	-.4782856504122416	-7.6124124526977540	

Part B. Simple Linear Regression

- Application 1: Check the linear relationship
- A. Compute Correlations

The screenshot shows the SPSS software interface. The 'Analyze' menu is open, and the 'Correlate' option is highlighted. The 'Bivariate...' option is also highlighted. Three callout boxes with arrows point to these options: '1. Click "Analyze" tab' points to the 'Analyze' menu, '2. Click "Correlate"' points to the 'Correlate' option, and '3. Click "Bivariate"' points to the 'Bivariate...' option.

1. Click "Analyze" tab

2. Click "Correlate"

3. Click "Bivariate"

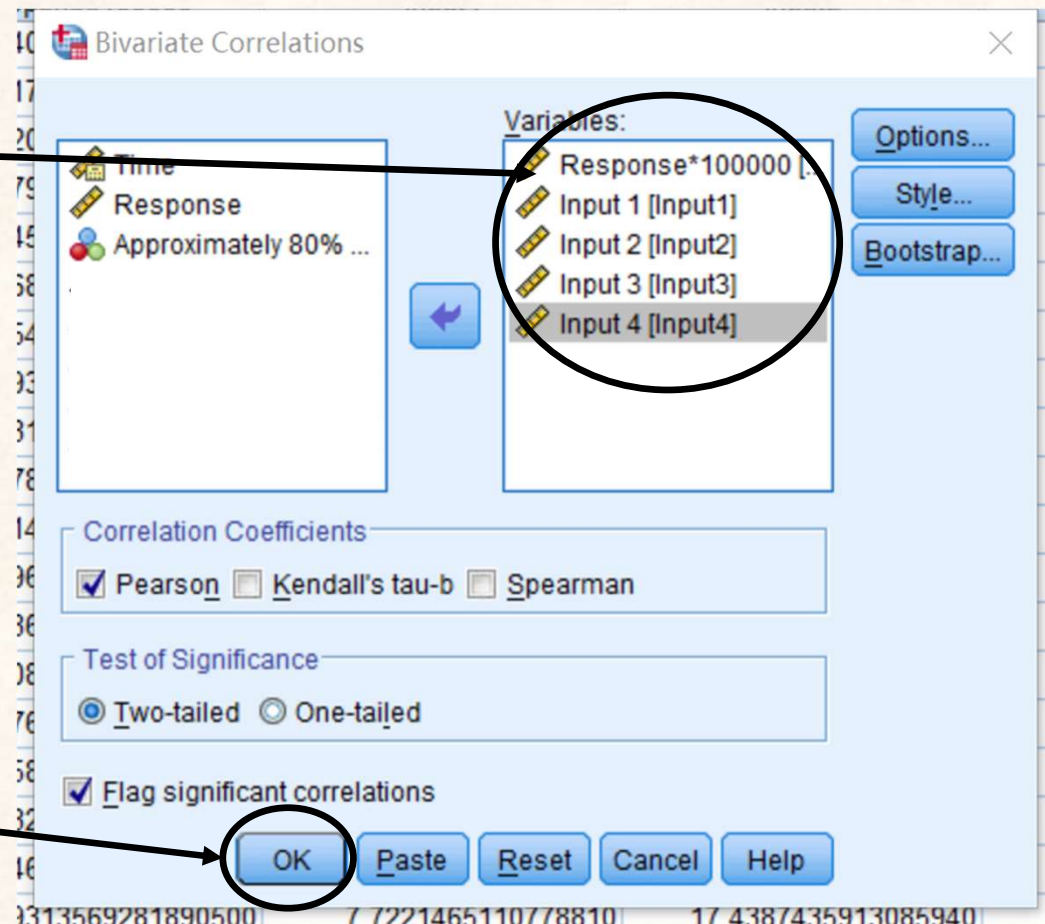
	Input1	Input2
7600	7.1575646400451660	17.3681430816650400
9860	9.2956762313842770	19.5979156494140620
9500	8.9122705459594730	19.5440826416015620
8500	8.59073486330	19.4902496337890620
8500	8.5002136230	18.3722763061523440
6400	8.7089443206787110	18.6906909942626950
8600	10.1771297454833980	19.9072399139404300
1260	10.1041145324707030	19.3486595153808600
5450	8.7552623748779300	18.6338386535644530
10060	8.6200866699218750	17.3237552642822270
9500	8.2210998535156250	17.1752586364746100
4400	8.6732320785522460	17.1253852844238280
7100	7.5828623771667480	17.3601074218750000
7100	8.0207948684692380	16.9656524658203120
5900	8.2677745819091800	17.1853485107421880
9800	7.7124676704406740	17.5111103057861330
10500	7.7221465110778810	17.4387435913085940
5560	8.4138078689575200	19.4264602661132800
10700	8.9186391830444340	19.6547069549560550

Part B. Simple Linear Regression

□ Application 1. A: Compute Correlations

1. Drag “**Response*100000**”, “**Input1**”, “**Input2**”, “**Input3**”, “**Input4**” from the left column to the right column

2. Click “**OK**”



Part B. Simple Linear Regression

□ Application 1. A: Compute Correlations

Output of Correlations

→ Correlations

		Correlations				
		Response*10 0000	Input 1	Input 2	Input 3	Input 4
Response*100000	Pearson Correlation	1	.356**	.485**	-.667**	.124**
	Sig. (2-tailed)		.000	.000	.000	.000
	N	48059	48059	48059	48059	48059
Input 1	Pearson Correlation	.356**	1	.869**	-.388**	.056**
	Sig. (2-tailed)	.000		.000	.000	.000
	N	48059	48059	48059	48059	48059
Input 2	Pearson Correlation	.485**	.869**	1	-.694**	.095**
	Sig. (2-tailed)	.000	.000		.000	.000
	N	48059	48059	48059	48059	48059
Input 3	Pearson Correlation	-.667**	-.388**	-.694**	1	-.136**
	Sig. (2-tailed)	.000	.000	.000		.000
	N	48059	48059	48059	48059	48059
Input 4	Pearson Correlation	.124**	.056**	.095**	-.136**	1
	Sig. (2-tailed)	.000	.000	.000	.000	
	N	48059	48059	48059	48059	48059

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

Part B. Simple Linear Regression

- Application 1: Check the linear relationship
- B. Observe the relationship in a scatter plot.

1. Click “**Graphs**” tab

2. Click “**Legacy Dialogs**”

3. Click “**Scatter/Dot**”

The screenshot shows the Minitab software interface. The 'Graphs' menu is open, and 'Legacy Dialogs' is selected. The 'Scatter/Dot...' option is highlighted in the submenu. The background shows a data table with columns: Time, Response, and Respo.

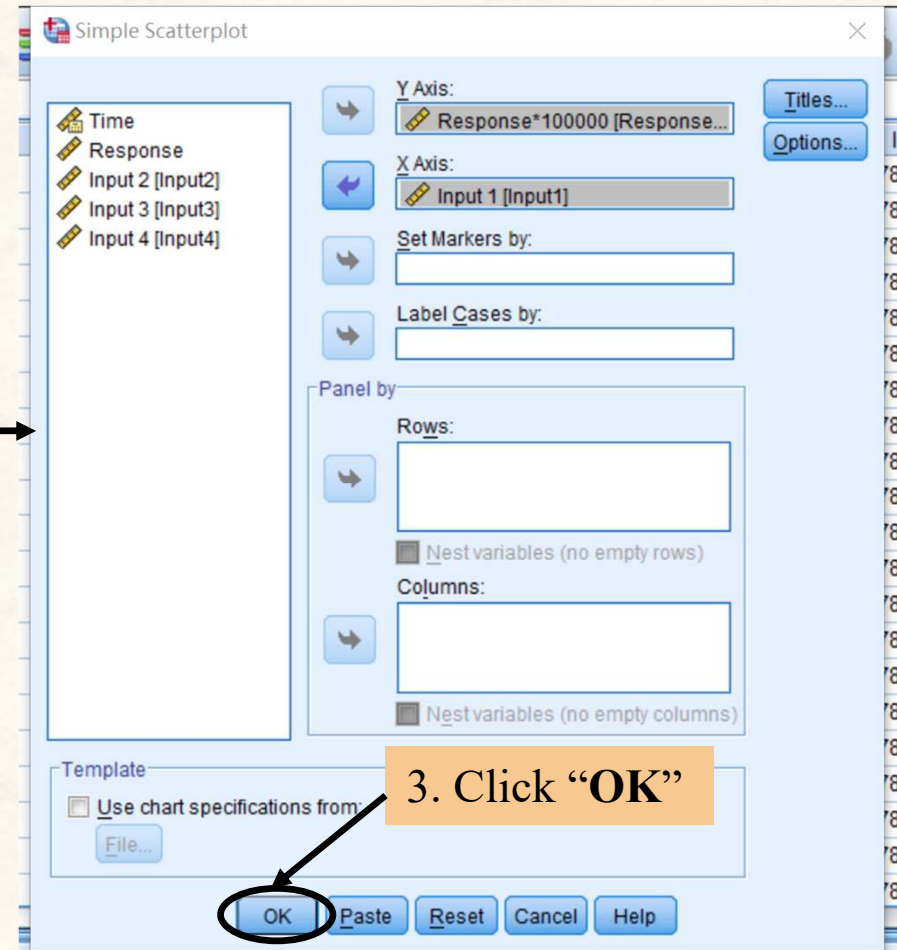
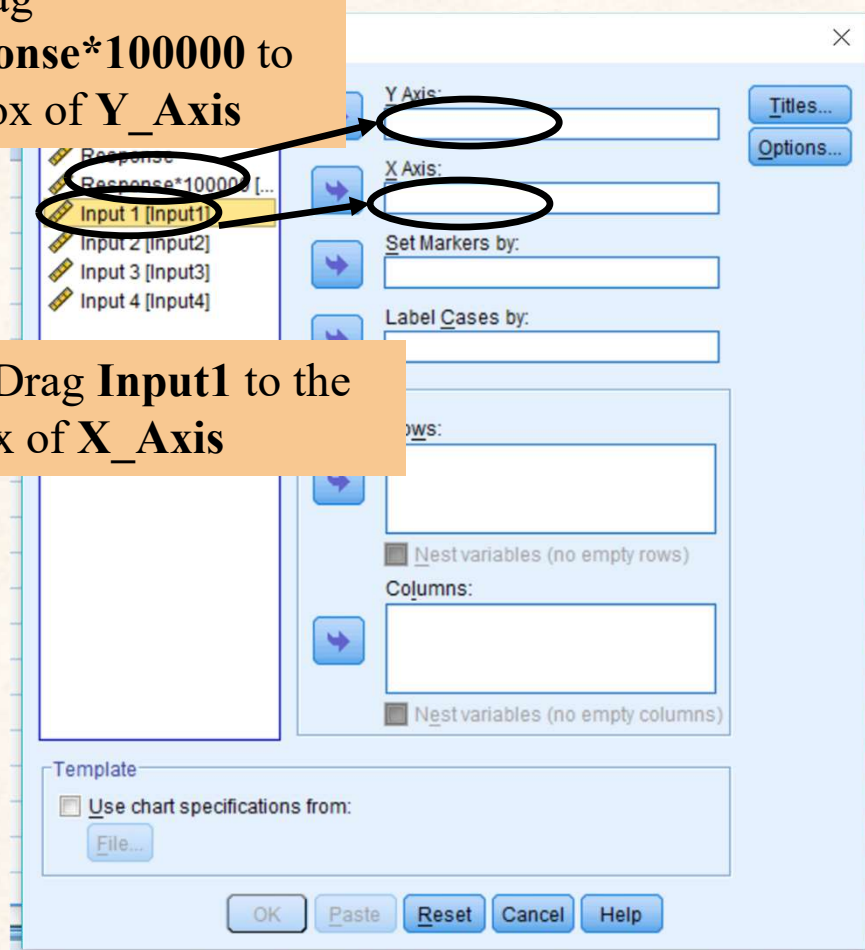
	Time	Response	Respo
1	01-Oct-2008	.0003535419957635	35.354
2	01-Oct-2008	.0003817622656483	38.176
3	01-Oct-2008	.0003710044730330	37.100
4	01-Oct-2008	.0003192898713623900	31.929
5	01-Oct-2008	.000425318686230933000	42.532
6	01-Oct-2008	.0003280453887398	32.804
7	01-Oct-2008	.0003090853467528	30.908
8	01-Oct-2008	.0003585741727826	35.857
9	01-Oct-2008	.0003358370583409	33.583
10	01-Oct-2008	.0003907306916921	39.073
11	01-Oct-2008	.0004415300483621	44.153
12	01-Oct-2008	.0004039063300824	40.390
13	01-Oct-2008	.0003848553583987	38.485
14	01-Oct-2008	.0003667757911329	36.677
15	01-Oct-2008	.0003441899034093	34.418
16	01-Oct-2008	.0002949460057972	29.494
17	01-Oct-2008	.0003022158733150	30.221
18	01-Oct-2008	.0003285147653682	32.851
19	01-Oct-2008	.0003629734263869	36.297
20	01-Oct-2008	.0004120962422249	41.209

Part B. Simple Linear Regression

- Application 1. B: Draw a scatter plot
- Take **Input1** and **Response*100000** as example

1. Drag
Response*100000 to
the box of **Y_Axis**

2. Drag **Input1** to the
box of **X_Axis**



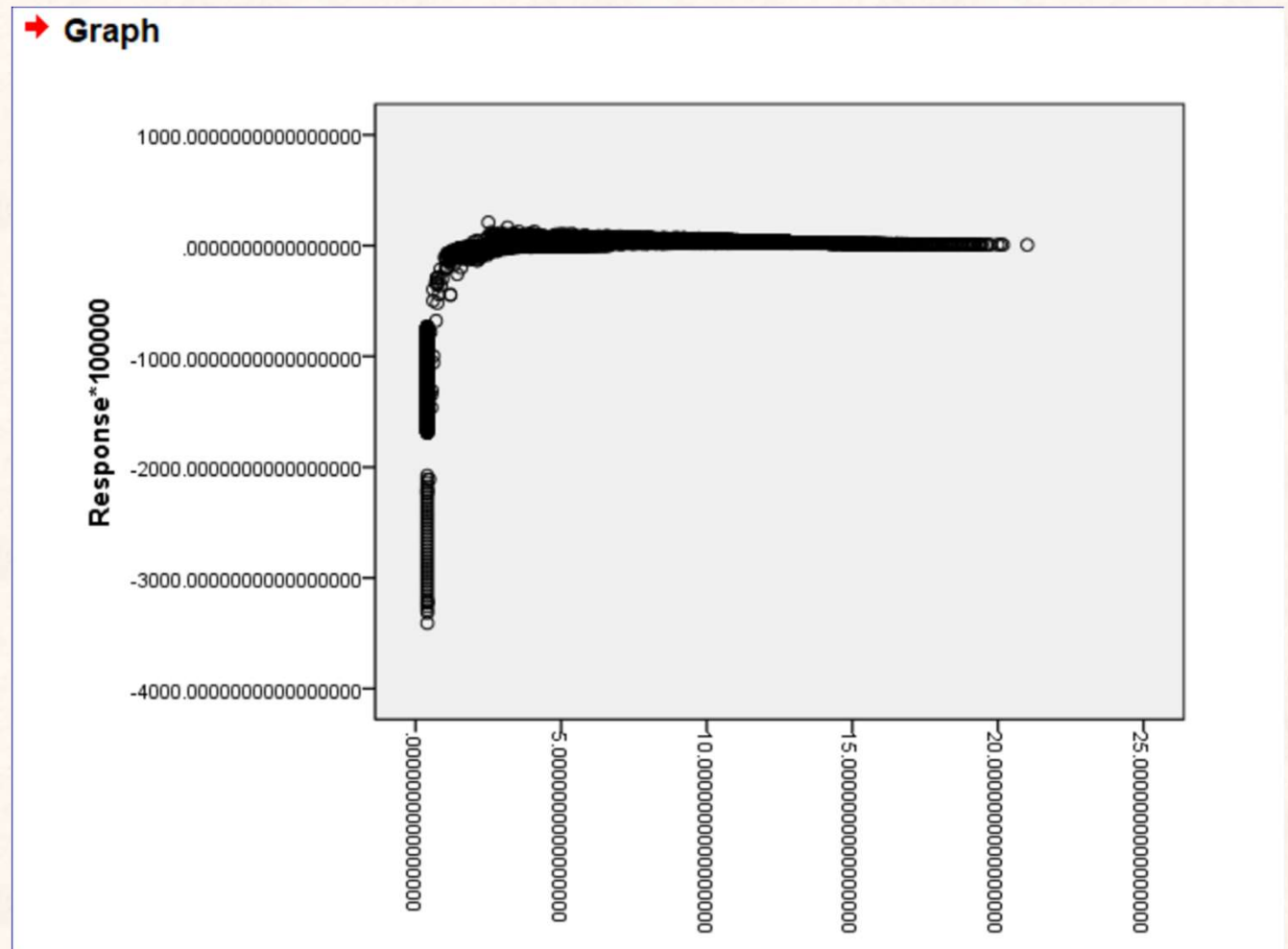
3. Click "OK"

Part B. Simple Linear Regression

- Application 1. B: Draw a scatter plot

Result:

It would be good to use linear regression if the shape looks like a line.



Part B. Simple Linear Regression

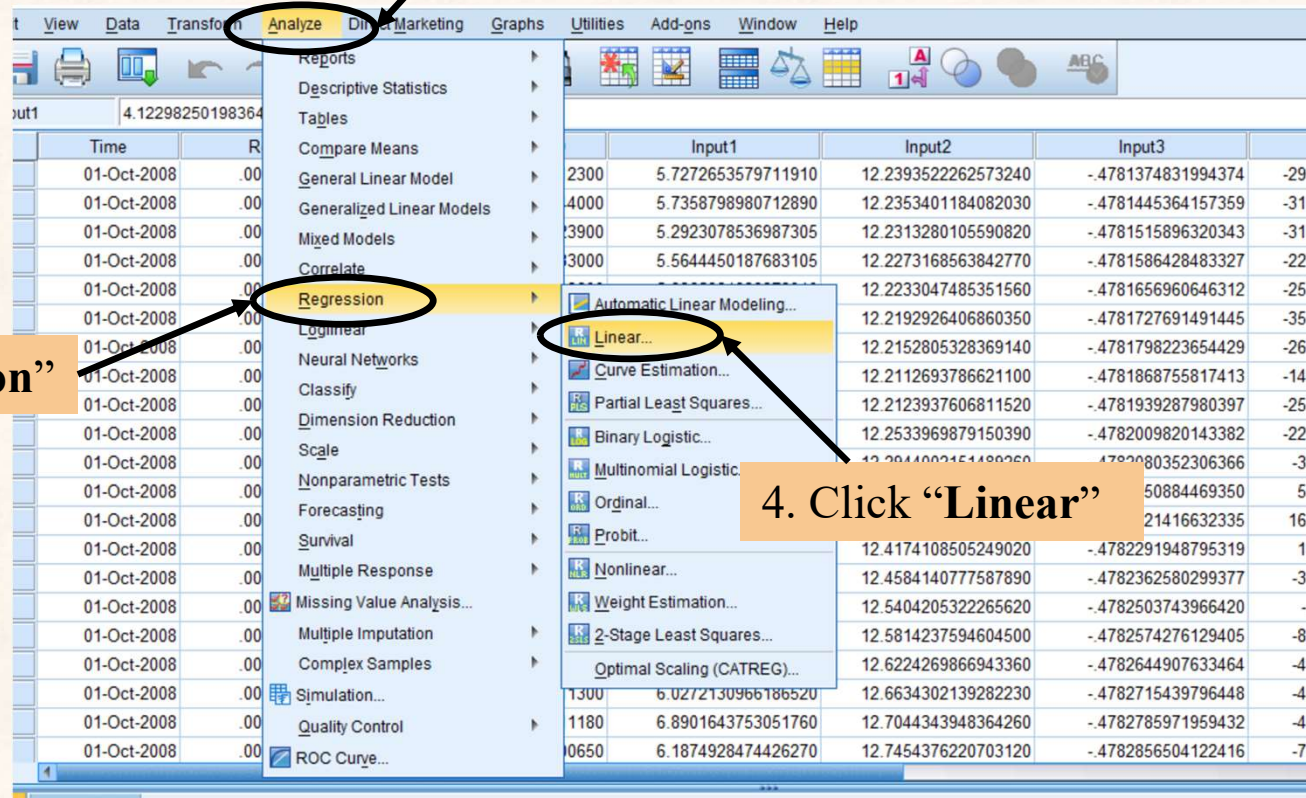
- Application 2: Conduct simple linear regression, using **Input1** as independent variable and **Response*100000** as dependent variable.

1. Open “Data_train”

2. Click “Analyze” tab

3. Click “Regression”

4. Click “Linear”

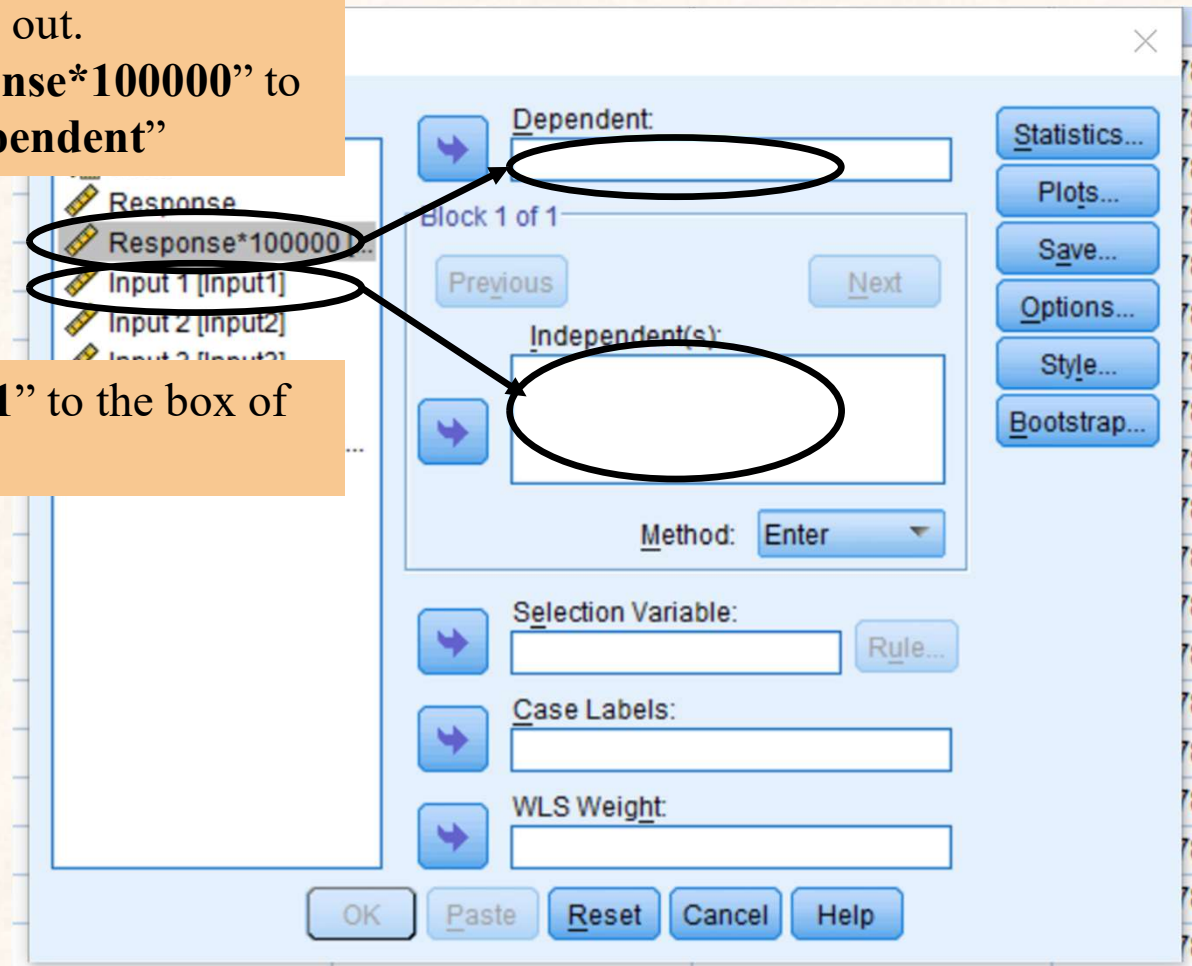


Part B. Simple Linear Regression

A window pops out.

1. Drag “**Response*100000**” to the box of “**Dependent**”

2. Drag “**Input1**” to the box of “**Independent**”



Part B. Simple Linear Regression

1. Click “Statistics...”

2. In the new window, we check on “Estimates”, “Confidence intervals” and “Model fit”

The image shows two overlapping software windows from a statistical package. The background window is titled "Linear Regression" and has a list of variables on the left: Time, Response, Input 1 [Input1], Input 2 [Input2], Input 3 [Input3], Input 4 [Input4], and Approximately 80% The "Dependent:" field contains "Response*100000 [Respon...]" and the "Independent(s):" field contains "Input 1 [Input1]". The "Method:" is set to "Enter". The "Statistics..." button is circled in black. The foreground window is titled "Linear Regression: Statistics" and has several sections. In the "Regression coefficients" section, the "Estimates", "Confidence intervals", and "Model fit" checkboxes are checked and circled in black. The "Level(%)" is set to 95. In the "Residuals" section, the "Outliers outside:" option is selected with a value of 3. The "Continue" button at the bottom is also circled in black. An arrow points from the "Continue" button to a text box at the bottom right.

3. Click “Continue”

Part B. Simple Linear Regression

1. Click “Plots...”

2. In the new window, we check on “Histogram” of Standardize Residual Plots

The image shows two overlapping SPSS dialog boxes. The 'Linear Regression' dialog box on the left has 'Dependent:' set to 'Response*100000 [Respon...]' and 'Independent(s):' set to 'Input 1 [Input1]'. The 'Plots...' button is circled. The 'Linear Regression Plots' dialog box on the right has 'DEPENDNT' set to '*ZPRED', '*ZRESID', '*DRESID', '*ADJPRED', '*SRESID', and '*SDRESID'. Under 'Standardized Residual Plots', the 'Histogram' checkbox is checked and circled. The 'Continue' button is also circled. Arrows point from the text instructions to these specific elements.

Linear Regression

Dependent: Response*100000 [Respon...]

Block 1 of 1

Previous Next

Independent(s): Input 1 [Input1]

Method: Enter

Selection Variable:

Case Labels:

WLS Weight:

OK Paste Reset Cancel Help

Statistics... Plots... Save... Options... Style... Bootstrap...

Linear Regression Plots

DEPENDNT

*ZPRED
*ZRESID
*DRESID
*ADJPRED
*SRESID
*SDRESID

Scatter 1 of 1

Previous Next

Y:

X:

Standardized Residual Plots

☒ Histogram
☐ Normal probability plot

☐ Produce all partial plots

Continue Cancel Help

3. Click “Continue”

Part B. Simple Linear Regression

1. Click "Save..."

2. In the new window, we check on "Unstandardized" of Predicted Values

3. Click "Continue"

4. Click "OK"

The image shows the SPSS Linear Regression dialog box and its 'Save Predicted Values' sub-dialog. The main dialog has 'Response*100000 [Respon...]' as the dependent variable and 'Input 1 [Input1]' as the independent variable. The 'Save...' button is circled. The sub-dialog shows 'Unstandardized' checked under 'Predicted Values'. The 'Continue' button is circled. The 'OK' button in the main dialog is also circled.

Time	Response	Input 1 [Input1]	Input 2 [Input2]	Input 3 [Input3]	Input 4 [Input4]	Approximately 80% ...
6.8901643753051760	12.7044343948364260	-4.782785971959432				
6.1874928474426270	12.7454376220703120	-4.782856504122416				

Part B. Simple Linear Regression

*Output2 [Document2] - IBM SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Output
Log
Regression
Title
Notes
Active Dataset
Variables Entered
Model Summary
ANOVA
Coefficients
Residuals Statistics

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Input 1 ^b	.	Enter

a. Dependent Variable: Response*100000
b. All requested variables entered.

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.356 ^a	.126	.126	177.5520380

a. Predictors: (Constant), Input 1
b. Dependent Variable: Response*100000

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	219376790.1	1	219376790.1	6958.880	.000 ^b
	Residual	1514983767	48057	31524.726		
	Total	1734360557	48058			

a. Dependent Variable: Response*100000
b. Predictors: (Constant), Input 1

R Square = 0.126

**Sig ≤ 0.05
Result is significant**

Part B. Simple Linear Regression

$\text{Sig} \leq 0.05$

The regression coefficients are significant.

Coefficients ^a							
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
1 (Constant)	-168.527	2.278		-73.972	.000	-172.993	-164.062
Input 1	23.765	.285	.356	83.420	.000	23.207	24.323

a. Dependent Variable: Response*100000

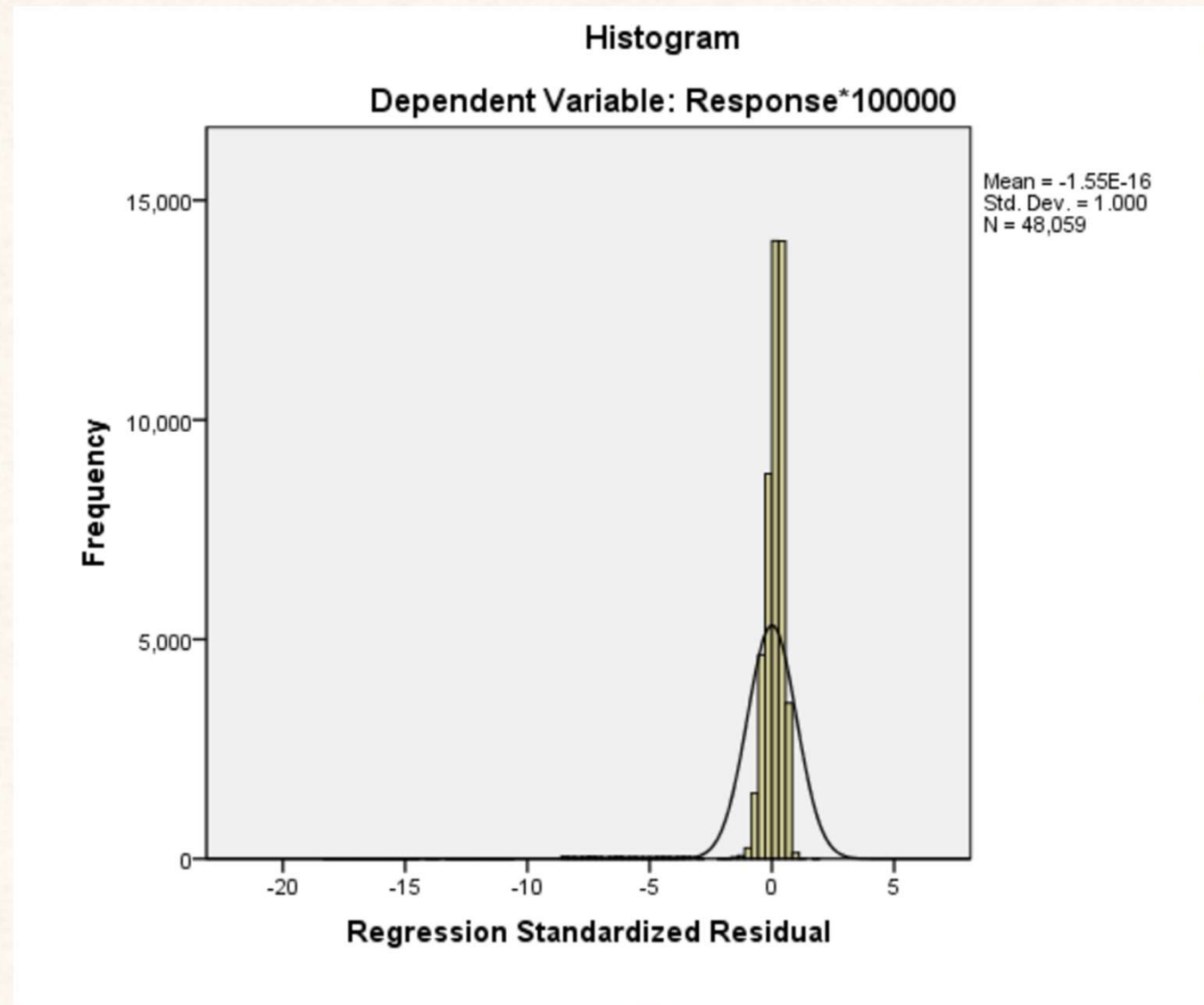
Regression coefficients

Residuals Statistics ^a					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-159.021133	330.7330322	9.110077238	67.56355491	48059
Residual	-3249.54614	318.6692505	.0000000000	177.5501907	48059
Std. Predicted Value	-2.488	4.760	.000	1.000	48059
Std. Residual	-18.302	1.795	.000	1.000	48059

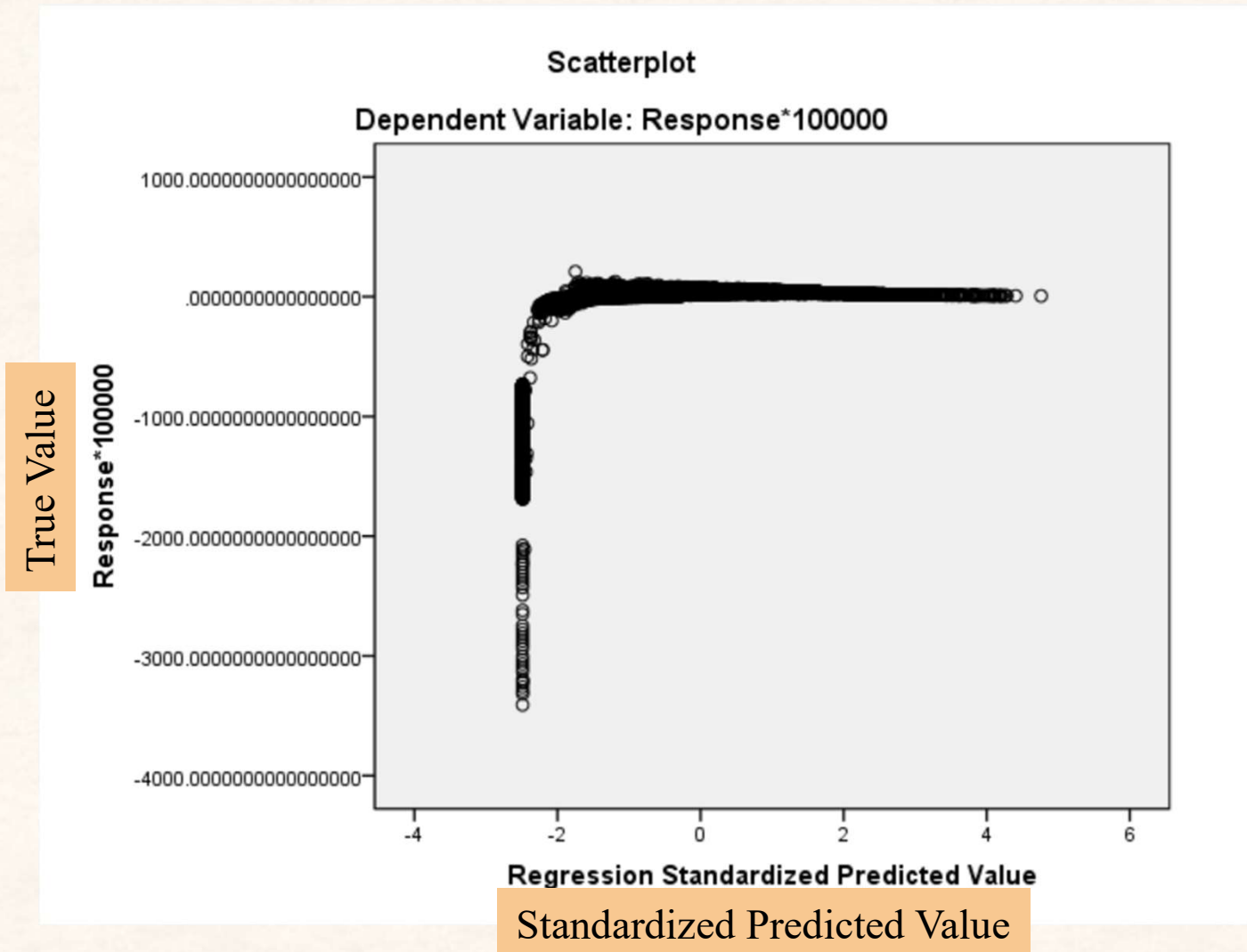
a. Dependent Variable: Response*100000

Part B. Simple Linear Regression

According to the assumption of the linear regression, residuals should be normally distributed. You can check the results by this plot.



Part B. Simple Linear Regression



Part B. Simple Linear Regression

*Untitled4 [Data_train] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

9 :

	ut1	Input2	Input3	Input4	filter_\$	PRE_1
1	53579711910	12.2393522262573240	-.4781374831994374	-29.3236083984375000	1	-32.41892
2	98980712890	12.2353401184082030	-.4781445364157359	-31.3377075195312500	1	-32.21420
3	078536987305	12.2313280105590820	-.4781515896320212	-31.1882410551112250	1	-42.75567
4	50187683105	12.2273168563842770	-.4781515896320212	-31.1882410551112250	1	-36.28834
5	321838378910	12.2233047485351560	-.4781515896320212	-31.1882410551112250	1	-29.82101
6	193489074710	12.2192926406860350	-.4781515896320212	-31.1882410551112250	1	-23.35368
7	316284179690	12.2152805328369140	-.4781515896320212	-31.1882410551112250	1	-18.62417
8	130279541020	12.2112693786621100	-.4781515896320212	-31.1882410551112250	1	-24.17372
9	55395507810	12.2123937606811520	-.4781515896320212	-31.1882410551112250	1	-19.29828
10	079475402830	12.2533969879150390	-.4782009820143382	-22.4881954193115230	1	-25.07216
11	786186218260	12.2944002151489260	-.4782080352306366	-3.8344509601593018	1	-29.32441
12	96563720700	12.3354034423828120	-.4782150884469350	5.9199233055114750	1	-23.91572
13	503440856930	12.3764066696167000	-.4782221416632335	16.8376522064209000	1	-22.24312
14	310317993160	12.4174108505249020	-.4782291948795319	1.2255277633666992	1	-20.57053
15	355621337890	12.4584140777587890	-.4782362580299377	-3.7134051322937010	1	-17.78755
16	01281738280	12.5404205322265620	-.4782503743966420	-.2330269664525986	1	-11.78298
17	28486633300	12.5814237594604500	-.4782574276129405	-8.3876247406005860	1	-15.39852
18	83064270020	12.6224269866943360	-.4782644907633464	-4.0000000000000000	1	-19.75047
19	130966186520	12.6634302139282230	-.4782715439796448	-4.0000000000000000	1	-25.29068
20	43753051760	12.7044343948364260	-.4782785971959432	-4.4861187934875490	1	-4.78268
21	28474426270	12.7454376220703120	-.4782856504122416	-7.6124124526977540	1	-21.48163

Data View Variable View

IBM SPSS Statistics

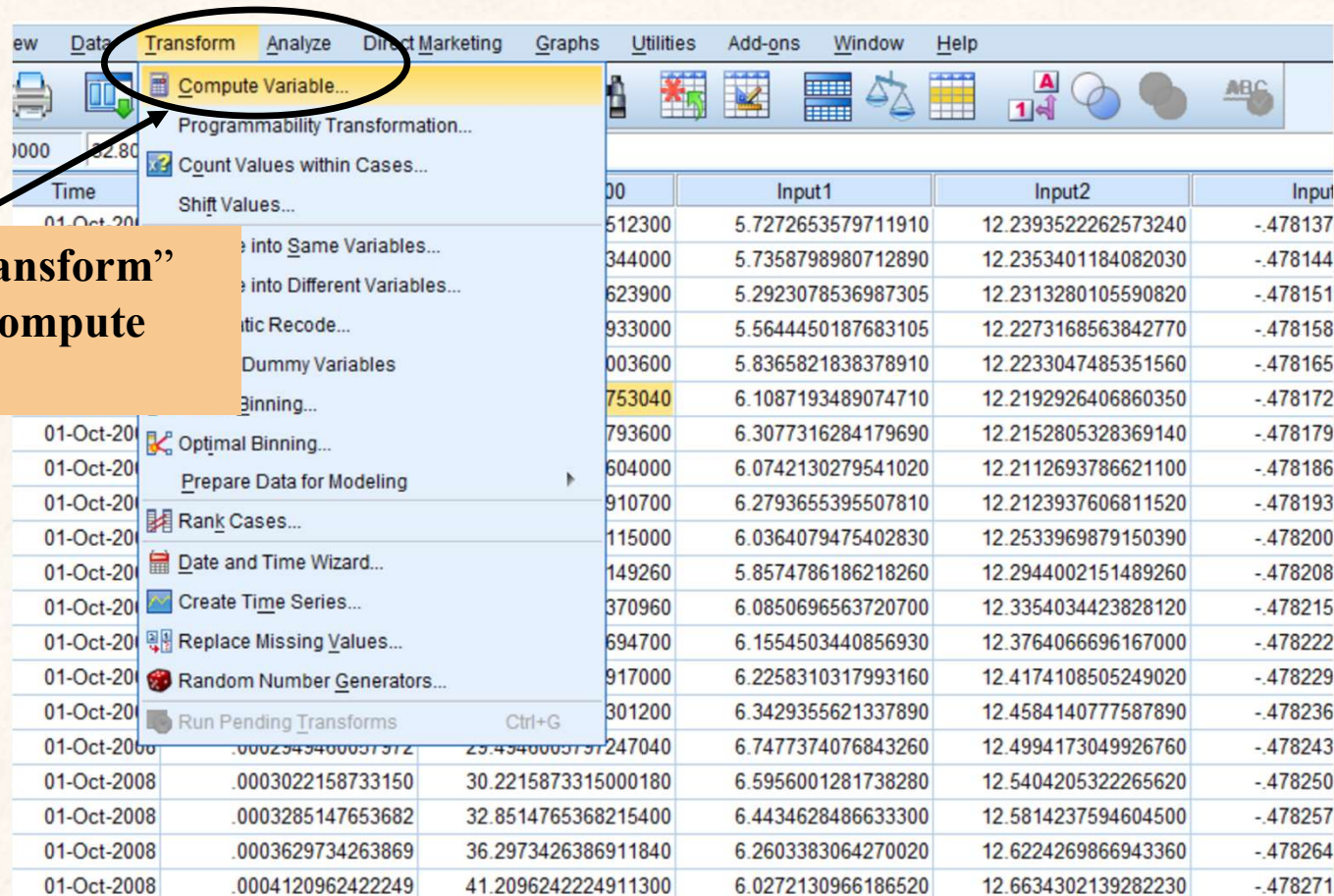
Move back to the Data_train, the predicted values are computed as Pre_1 at the far right side.

Part B. Simple Linear Regression

- Make prediction based on the current model for test set

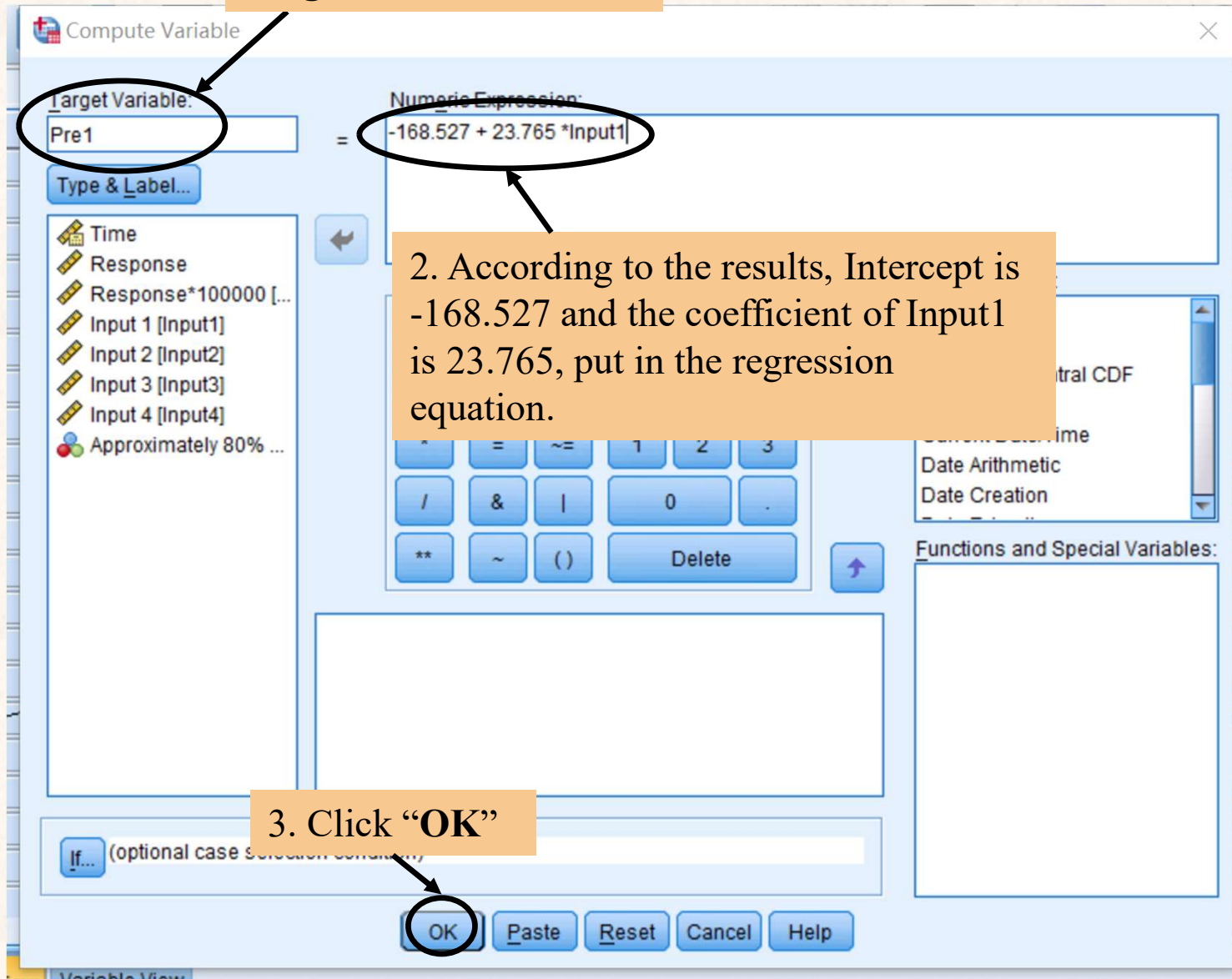
1. Move to “Data_test”

2. Click the “Transform” tab and click “Compute Variable”

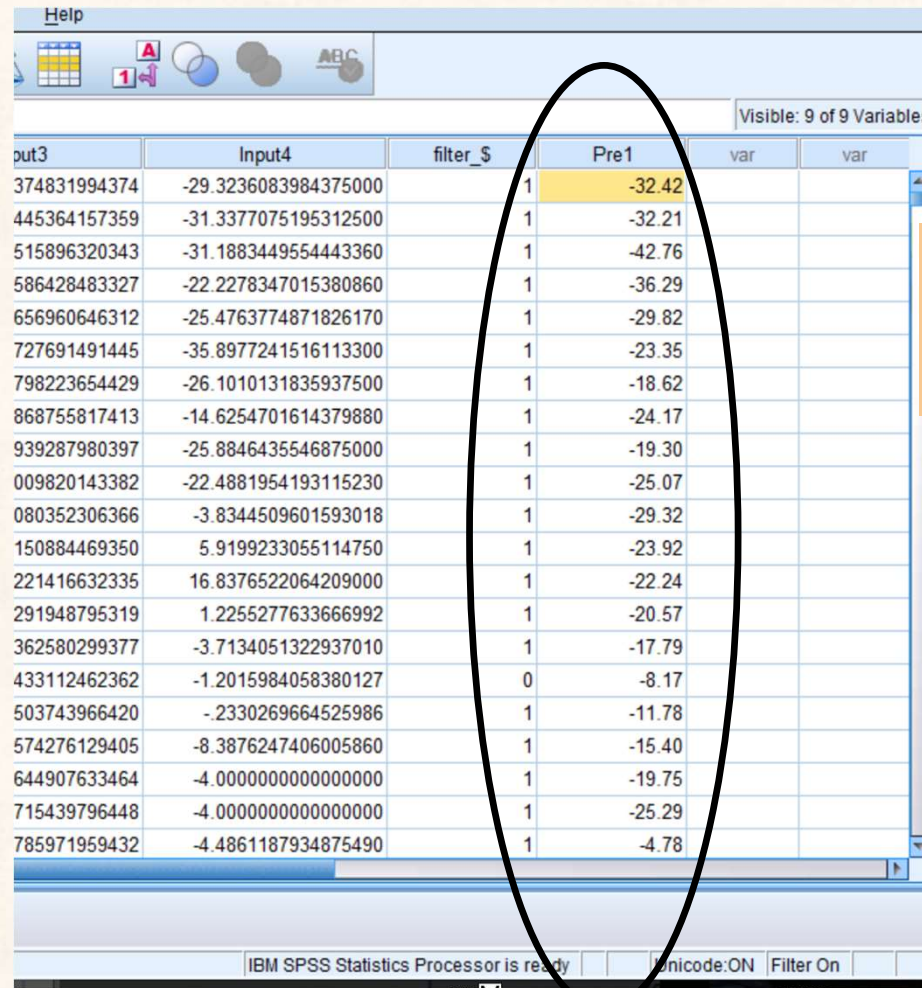


Part B. Simple Linear Regression

1. We put in **Pre1** into **Target Variable**.



Part B. Simple Linear Regression



Help

Visible: 9 of 9 Variables

out3	Input4	filter_\$	Pre1	var	var
374831994374	-29.3236083984375000	1	-32.42		
445364157359	-31.3377075195312500	1	-32.21		
515896320343	-31.1883449554443360	1	-42.76		
586428483327	-22.2278347015380860	1	-36.29		
656960646312	-25.4763774871826170	1	-29.82		
727691491445	-35.8977241516113300	1	-23.35		
798223654429	-26.1010131835937500	1	-18.62		
868755817413	-14.6254701614379880	1	-24.17		
939287980397	-25.8846435546875000	1	-19.30		
009820143382	-22.4881954193115230	1	-25.07		
080352306366	-3.8344509601593018	1	-29.32		
150884469350	5.9199233055114750	1	-23.92		
221416632335	16.8376522064209000	1	-22.24		
291948795319	1.2255277633666992	1	-20.57		
362580299377	-3.7134051322937010	1	-17.79		
433112462362	-1.2015984058380127	0	-8.17		
503743966420	-.2330269664525986	1	-11.78		
574276129405	-8.3876247406005860	1	-15.40		
644907633464	-4.0000000000000000	1	-19.75		
715439796448	-4.0000000000000000	1	-25.29		
785971959432	-4.4861187934875490	1	-4.78		

IBM SPSS Statistics Processor is ready Unicode: ON Filter On

The predicted values are computed as **Pre1** at the far right side.

Part B. Simple Linear Regression

*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor

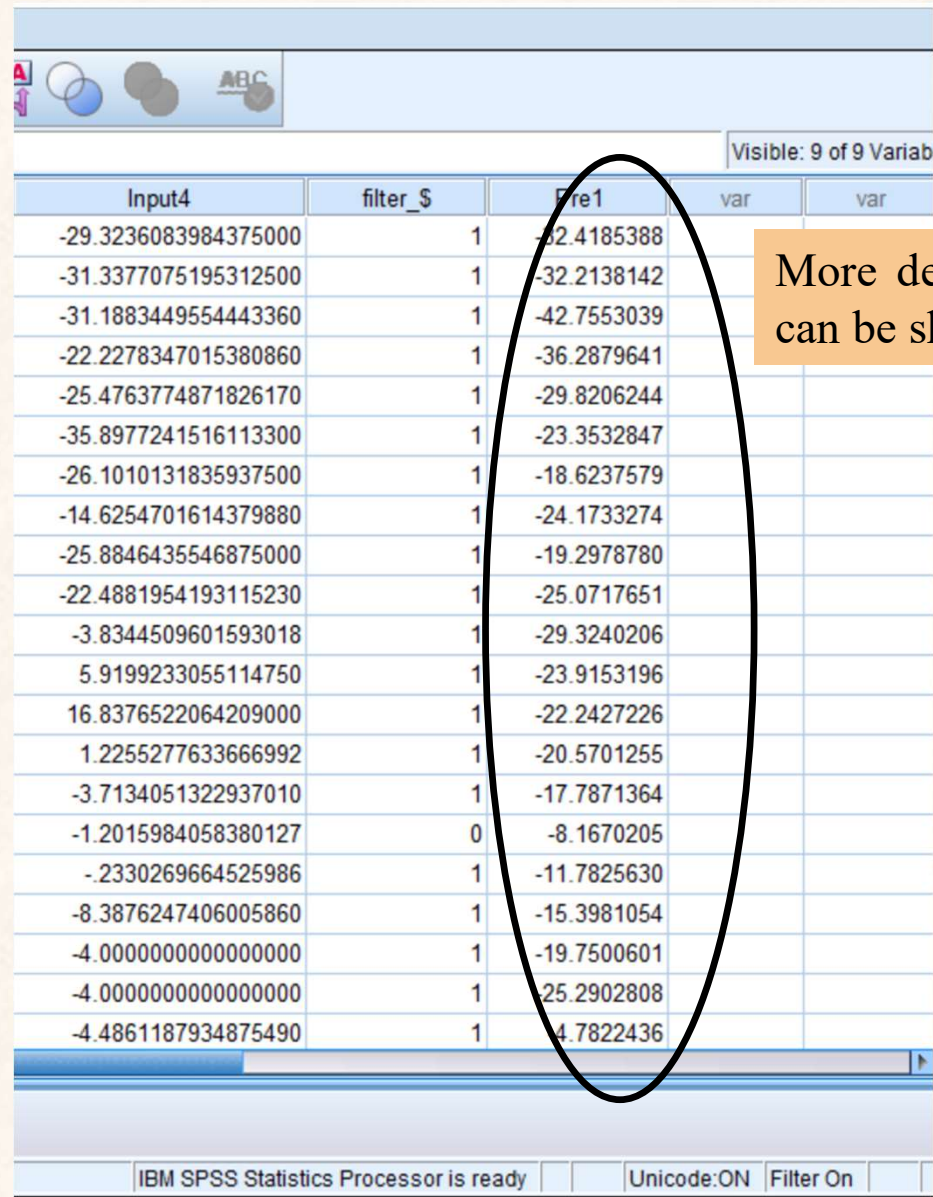
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Time	Date	11	0		None	None	11	Right	Scale	Input
2	Response	Numeric	17	16		None	None	17	Right	Scale	Input
3	Response1...	Numeric	17	16	Response*1000...	None	None	17	Right	Scale	Input
4	Input1	Numeric	17	16	Input 1	None	None	17	Right	Scale	Input
5	Input2	Numeric	17	16	Input 2	None	None	17	Right	Scale	Input
6	Input3	Numeric	17	16	Input 3	None	None	17	Right	Scale	Input
7	Input4	Numeric	17	16	Input 4	None	None	17	Right	Scale	Input
8	filter_\$	Numeric	1	0	Approximately ...	None	None	10	Right	Nominal	Input
9	Pre1	Numeric	8	7		None	None	10	Right	Scale	Input
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											

1. Click "Variable View"

2. Set this decimals to be greater, like 7.
So, put in 7 in this cell.

3. Click "Data View" to move back.

Part B. Simple Linear Regression



Visible: 9 of 9 Variables

Input4	filter_\$	Pre1	var	var
-29.3236083984375000	1	-32.4185388		
-31.3377075195312500	1	-32.2138142		
-31.1883449554443360	1	-42.7553039		
-22.2278347015380860	1	-36.2879641		
-25.4763774871826170	1	-29.8206244		
-35.8977241516113300	1	-23.3532847		
-26.1010131835937500	1	-18.6237579		
-14.6254701614379880	1	-24.1733274		
-25.8846435546875000	1	-19.2978780		
-22.4881954193115230	1	-25.0717651		
-3.8344509601593018	1	-29.3240206		
5.9199233055114750	1	-23.9153196		
16.8376522064209000	1	-22.2427226		
1.2255277633666992	1	-20.5701255		
-3.7134051322937010	1	-17.7871364		
-1.2015984058380127	0	-8.1670205		
-.2330269664525986	1	-11.7825630		
-8.3876247406005860	1	-15.3981054		
-4.0000000000000000	1	-19.7500601		
-4.0000000000000000	1	25.2902808		
-4.4861187934875490	1	4.7822436		

IBM SPSS Statistics Processor is ready Unicode:ON Filter On

More decimal places
can be shown.

Part B. Simple Linear Regression

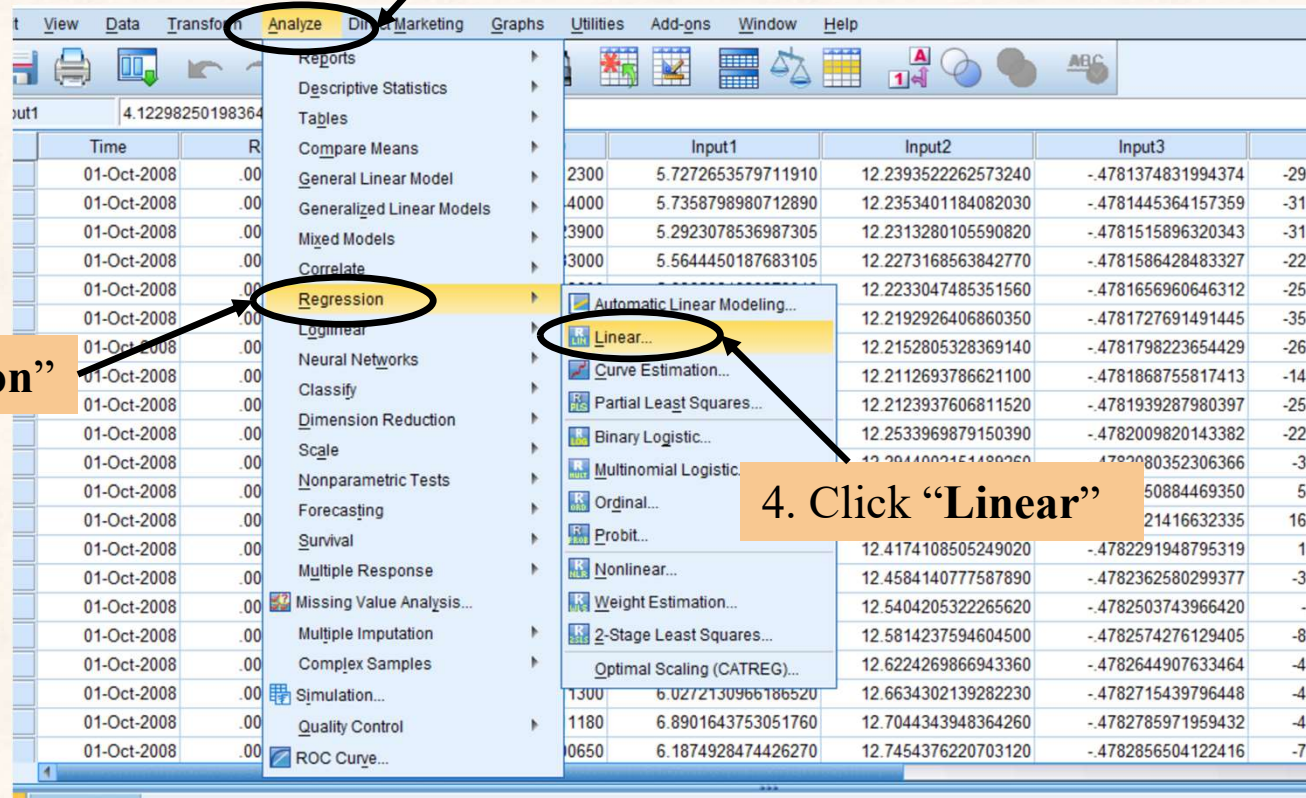
- Application 3: Conduct simple linear regression, using **Input2** as independent variable and **Response*100000** as dependent variable.

1. Open “Data_train”

2. Click “Analyze” tab

3. Click “Regression”

4. Click “Linear”



Part B. Simple Linear Regression

2. Drag the **Input2** to the box of Independent

1. Drag back **Input1** to the left column

3. Then click “OK”

The image displays two screenshots of the SPSS Linear Regression dialog box, illustrating the steps to set up a simple linear regression model.

Left Screenshot (Initial State):

- Dependent:** Response*100000 [Respon...]
- Independent(s):** Input 1 [Input1]
- Left Column (Available Variables):** Time, Response, Input 1 [Input1], Input 2 [Input2], Input 3 [Input3], Input 4 [Input4], Approximately 80% ..., Unstandardized Pre...

Right Screenshot (Final State):

- Dependent:** Response*100000 [Respon...]
- Independent(s):** Input 2 [Input2]
- Left Column (Available Variables):** Time, Response, Input 1 [Input1], Input 2 [Input2], Input 3 [Input3], Input 4 [Input4], Approximately 80% ..., Unstandardized Pre...

The steps are summarized as follows:

- Drag back **Input1** to the left column.
- Drag the **Input2** to the box of Independent.
- Then click “OK”.

Part B. Simple Linear Regression

→ Regression

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Input 2 ^b	.	Enter

a. Dependent Variable: Response*100000

b. All requested variables entered.

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.485 ^a	.235	.235	166.1675083

a. Predictors: (Constant), Input 2

b. Dependent Variable: Response*100000

R Square = 0.235

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	407427933.7	1	407427933.7	14755.658	.000 ^b
	Residual	1326932623	48057	27611.641		
	Total	1734360557	48058			

a. Dependent Variable: Response*100000

Sig ≤ 0.05

Result is significant

Part B. Simple Linear Regression

$\text{Sig} \leq 0.05$

The regression coefficients are significant.

Coefficients ^a							
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
1 (Constant)	-312.114	2.751		-113.459	.000	-317.506	-306.722
Input 2	21.348	.176	.485	121.473	.000	21.004	21.693

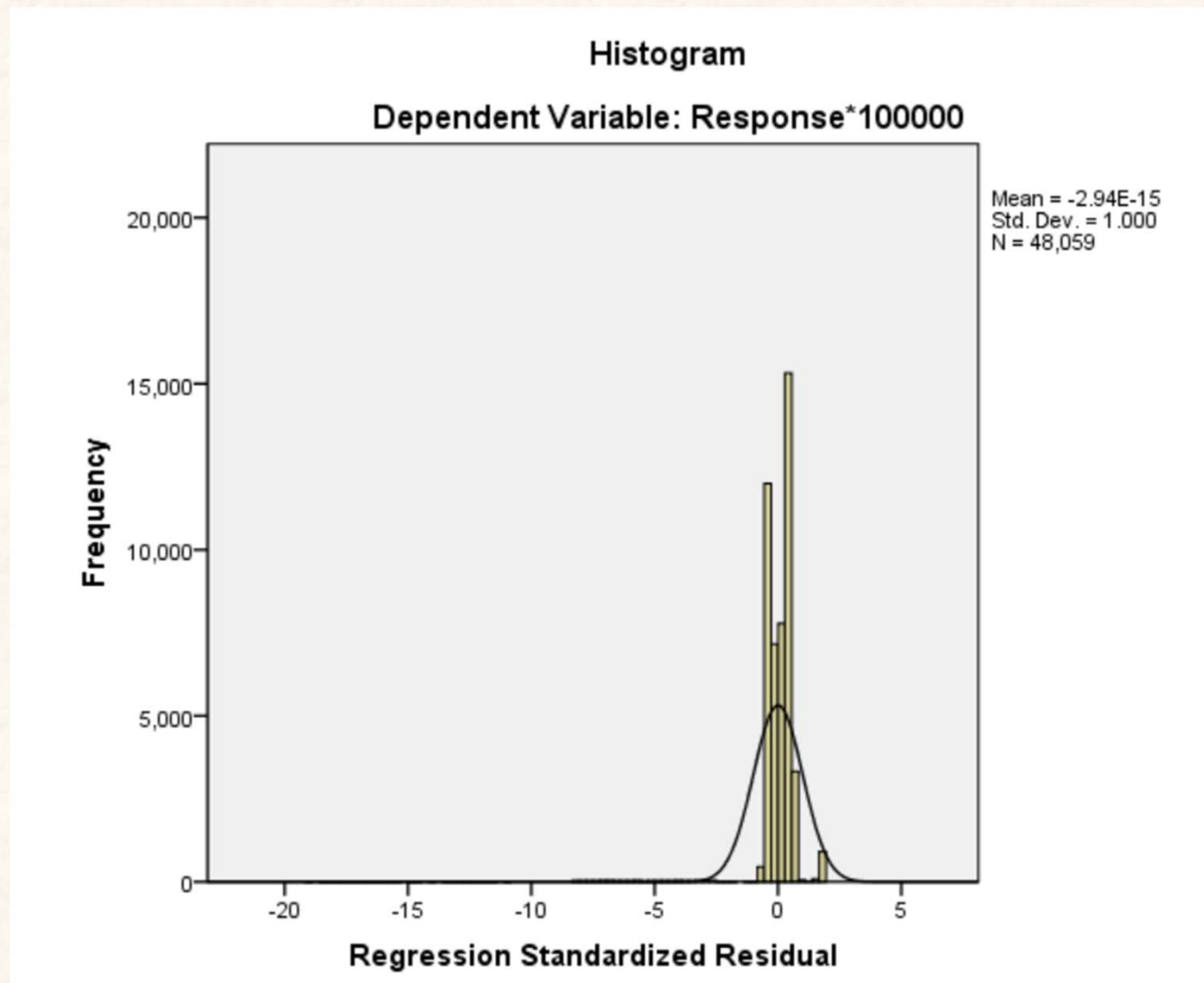
a. Dependent Variable: Response*100000

Regression coefficients

Residuals Statistics ^a					
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-312.114166	128.2289276	9.110077238	92.07517525	48059
Residual	-3156.21338	311.7413330	.0000000000	166.1657795	48059
Std. Predicted Value	-3.489	1.294	.000	1.000	48059
Std. Residual	-18.994	1.876	.000	1.000	48059

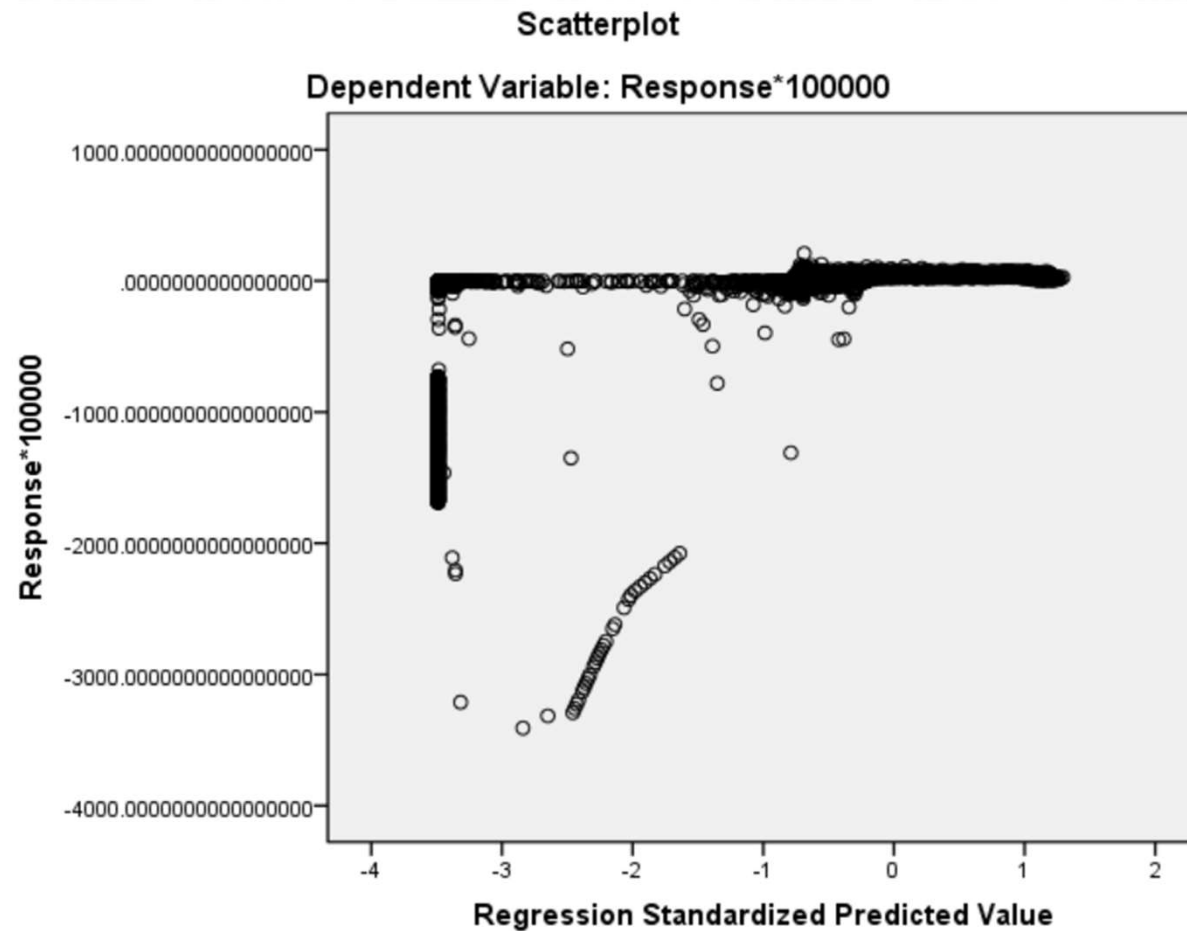
a. Dependent Variable: Response*100000

Part B. Simple Linear Regression



Part B. Simple Linear Regression

True Value



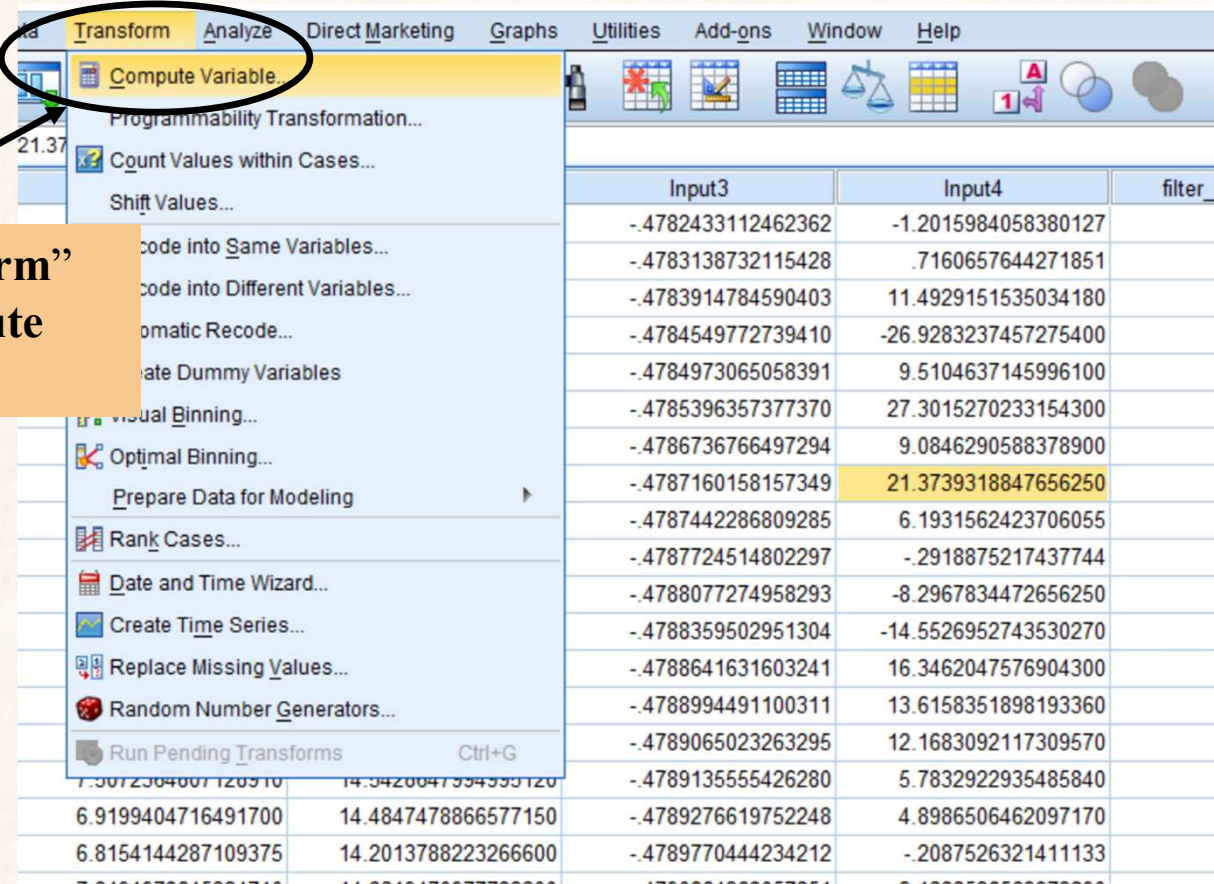
Standardized Predicted Value

Part B. Simple Linear Regression

- Make prediction based on the current model for test set

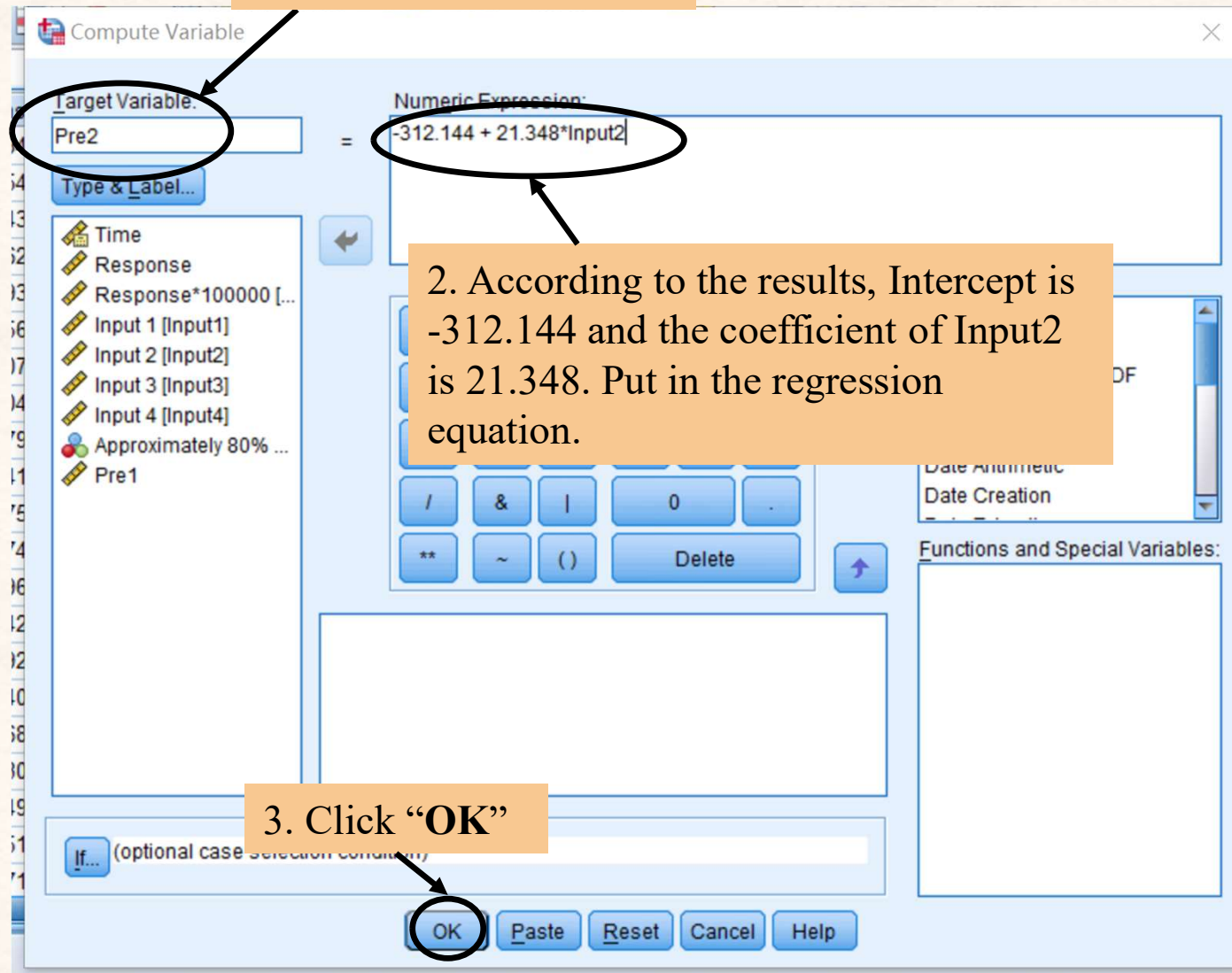
1. Move to “Data_test”

2. Click the “Transform” tab and click “Compute Variable”

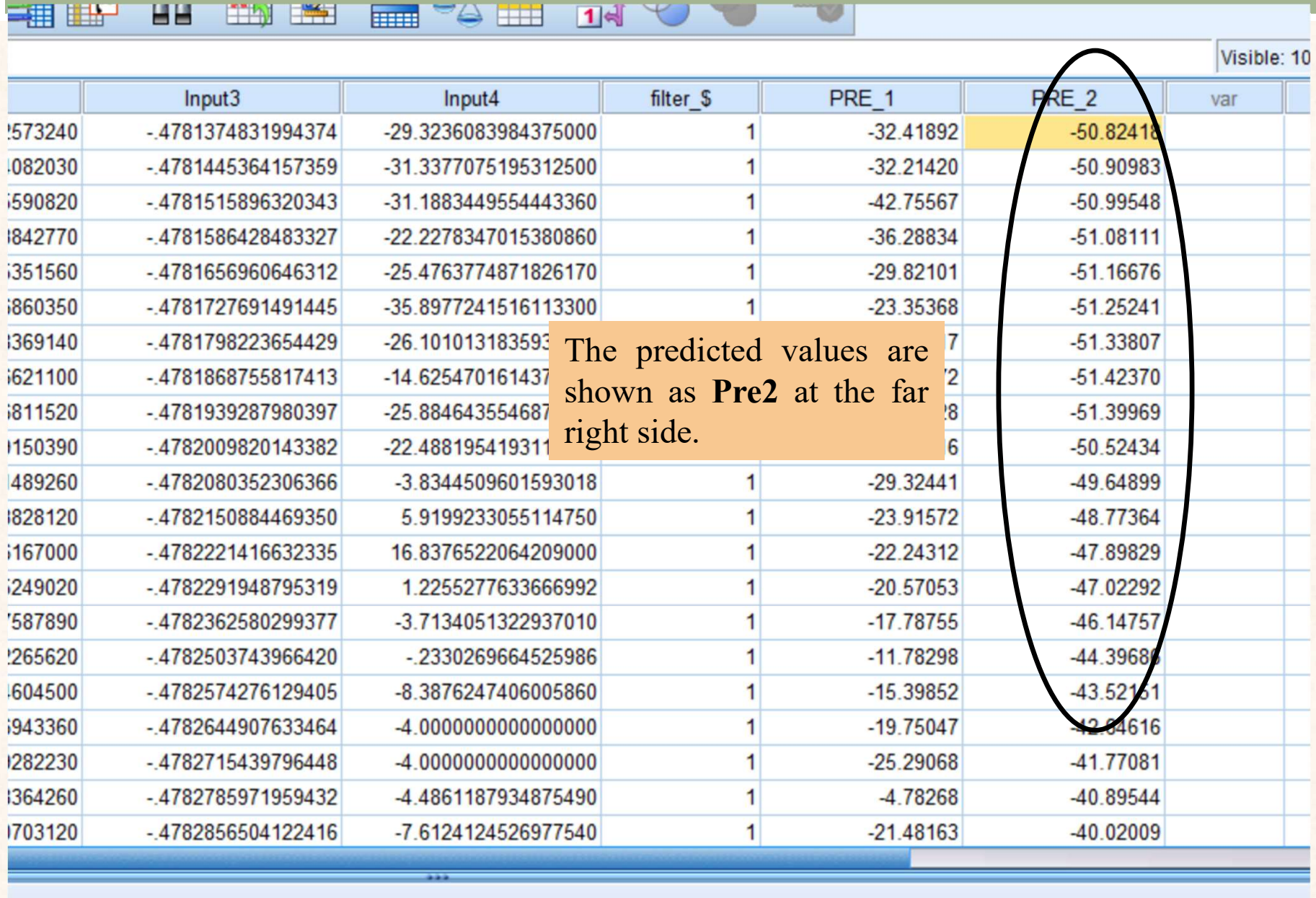


Part B. Simple Linear Regression

1. We put in Pre2 as the Target Variable.



Part B. Simple Linear Regression



Visible: 10

	Input3	Input4	filter_\$	PRE_1	PRE_2	var
573240	-.4781374831994374	-29.3236083984375000	1	-32.41892	-50.82418	
082030	-.4781445364157359	-31.3377075195312500	1	-32.21420	-50.90983	
590820	-.4781515896320343	-31.1883449554443360	1	-42.75567	-50.99548	
842770	-.4781586428483327	-22.2278347015380860	1	-36.28834	-51.08111	
351560	-.4781656960646312	-25.4763774871826170	1	-29.82101	-51.16676	
860350	-.4781727691491445	-35.8977241516113300	1	-23.35368	-51.25241	
369140	-.4781798223654429	-26.101013183593			-51.33807	
621100	-.4781868755817413	-14.625470161437			-51.42370	
811520	-.4781939287980397	-25.884643554687			-51.39969	
1150390	-.4782009820143382	-22.488195419311			-50.52434	
489260	-.4782080352306366	-3.8344509601593018	1	-29.32441	-49.64899	
828120	-.4782150884469350	5.9199233055114750	1	-23.91572	-48.77364	
167000	-.4782221416632335	16.8376522064209000	1	-22.24312	-47.89829	
249020	-.4782291948795319	1.2255277633666992	1	-20.57053	-47.02292	
587890	-.4782362580299377	-3.7134051322937010	1	-17.78755	-46.14757	
265620	-.4782503743966420	-.2330269664525986	1	-11.78298	-44.39686	
604500	-.4782574276129405	-8.3876247406005860	1	-15.39852	-43.52151	
943360	-.4782644907633464	-4.0000000000000000	1	-19.75047	-42.84616	
282230	-.4782715439796448	-4.0000000000000000	1	-25.29068	-41.77081	
364260	-.4782785971959432	-4.4861187934875490	1	-4.78268	-40.89544	
703120	-.4782856504122416	-7.6124124526977540	1	-21.48163	-40.02009	

Part C. Multiple Linear Regression

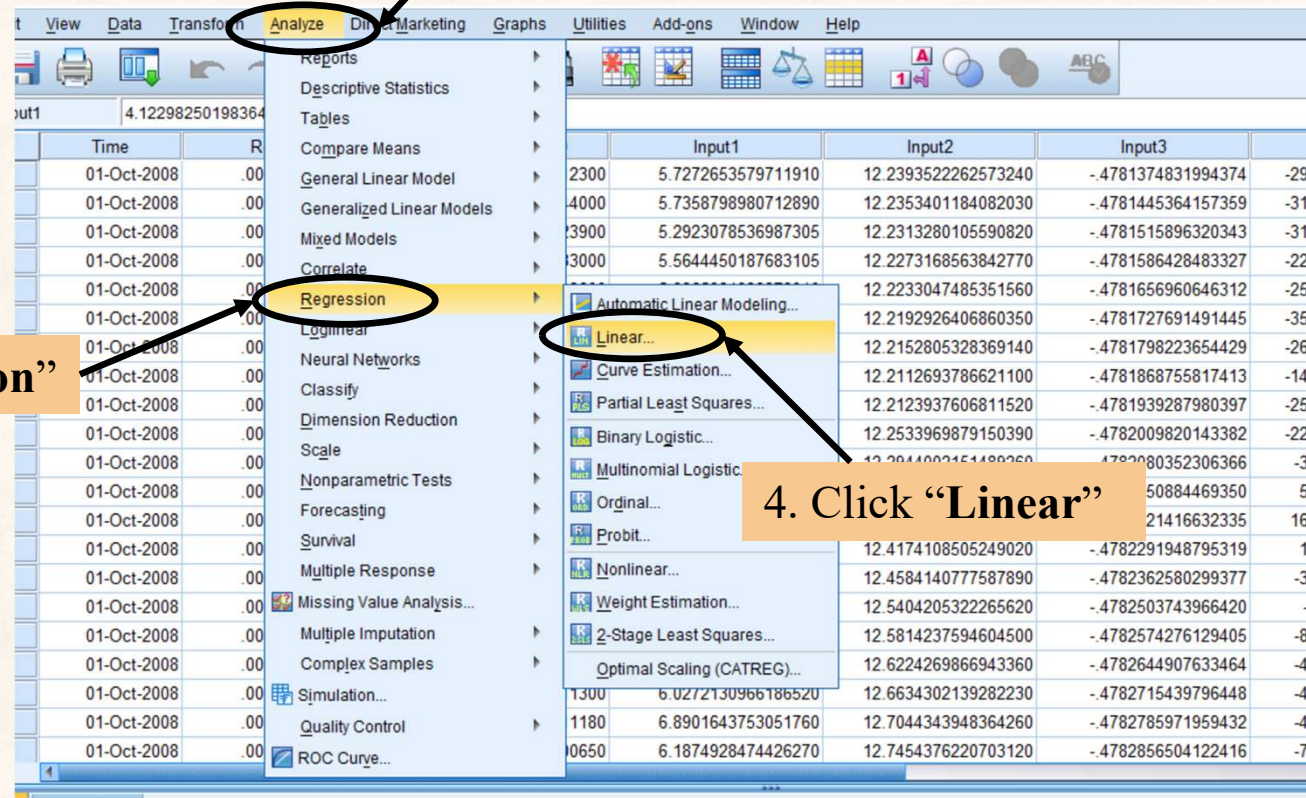
- Application 4: Conduct simple linear regression, using **Input1** and **Input2** as independent variables and **Response*100000** as dependent variable.

1. Open “Data_train”

2. Click “Analyze” tab

3. Click “Regression”

4. Click “Linear”



Part C. Multiple Linear Regression

The image shows a screenshot of the SPSS Linear Regression dialog box and its Statistics sub-dialog box. The main dialog box has a list of variables on the left, including 'Time', 'Response', 'Input 1 [Input1]', 'Input 2 [Input2]', 'Input 3 [Input3]', 'Input 4 [Input4]', 'Approximately 80% ...', 'Unstandardized Pre...', and 'Unstandardized Pre...'. The 'Dependent' variable is 'Response*100000 [Respon...]' and the 'Independent(s)' are 'Input 1 [Input1]' and 'Input 2 [Input2]'. The 'Method' is 'Enter'. The 'Statistics' sub-dialog box is open, showing options for 'Regression Coefficients' (Estimates, Confidence intervals, Level(%): 95, Covariance matrix) and 'Residuals' (Durbin-Watson, Casewise diagnostics, Outliers outside: 3 standard deviations, All cases). The 'Collinearity diagnostics' checkbox is checked. The 'Continue' button is highlighted.

1. Drag **Input1** and **Input2** to the box of Independents

2. Click “Statistics”

3. Also check on “Collinearity diagnostics”

4. Click “Continue”

5. Click “OK”

Part C. Multiple Linear Regression

[Data_train]

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Input 2, Input 1 ^b	.	Enter

a. Dependent Variable: Response*100000

b. All requested variables entered.

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.503 ^a	.253	.253	164.2437162

a. Predictors: (Constant), Input 2, Input 1

b. Dependent Variable: Response*100000

Adjusted R Square = 0.253

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	438001981.6	2	219000990.8	8118.365	.000 ^b
	Residual	1296358575	48056	26975.998		
	Total	1734360557	48058			

Sig ≤ 0.05
Result is significant

Part C. Multiple Linear Regression

Regression coefficients

$\text{Sig} \leq 0.05$

The regression coefficients are significant.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	-332.688	2.787		-119.376	.000	-338.151	-327.226		
	Input 1	-17.944	.533	-.269	-33.666	.000	-18.989	-16.900	.244	4.091
	Input 2	31.630	.351	.718	90.025	.000	30.941	32.318	.244	4.091

a. Dependent Variable: Response*100000

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions		
				(Constant)	Input 1	Input 2
1	1	2.920	1.000	.01	.00	.00
	2	.068	6.535	.65	.16	.01
	3	.012	15.765	.34	.84	.99

a. Dependent Variable: Response*100000

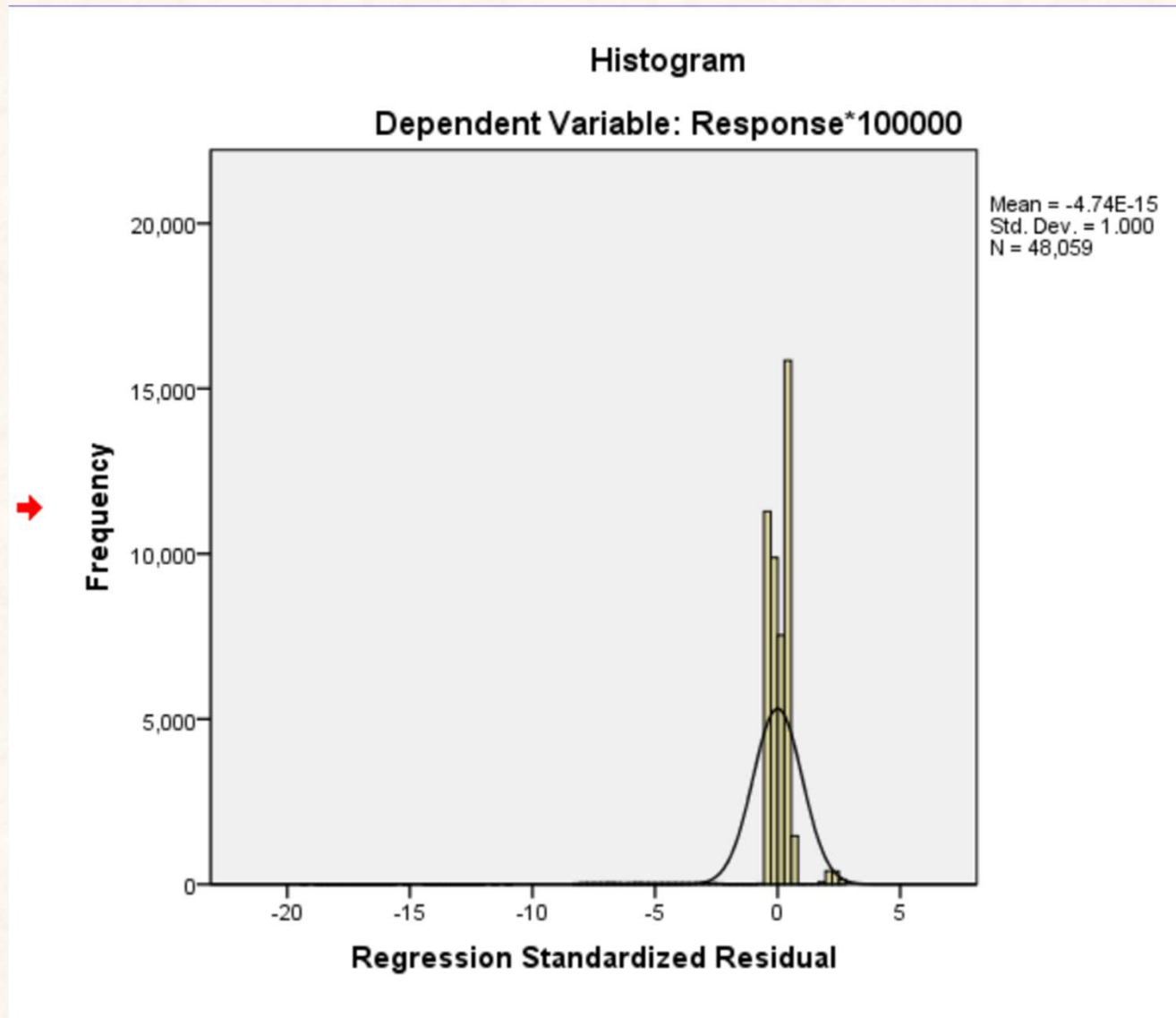
$\text{VIF} = 4.091 \leq 5$

There is very likely to be multicollinearity.

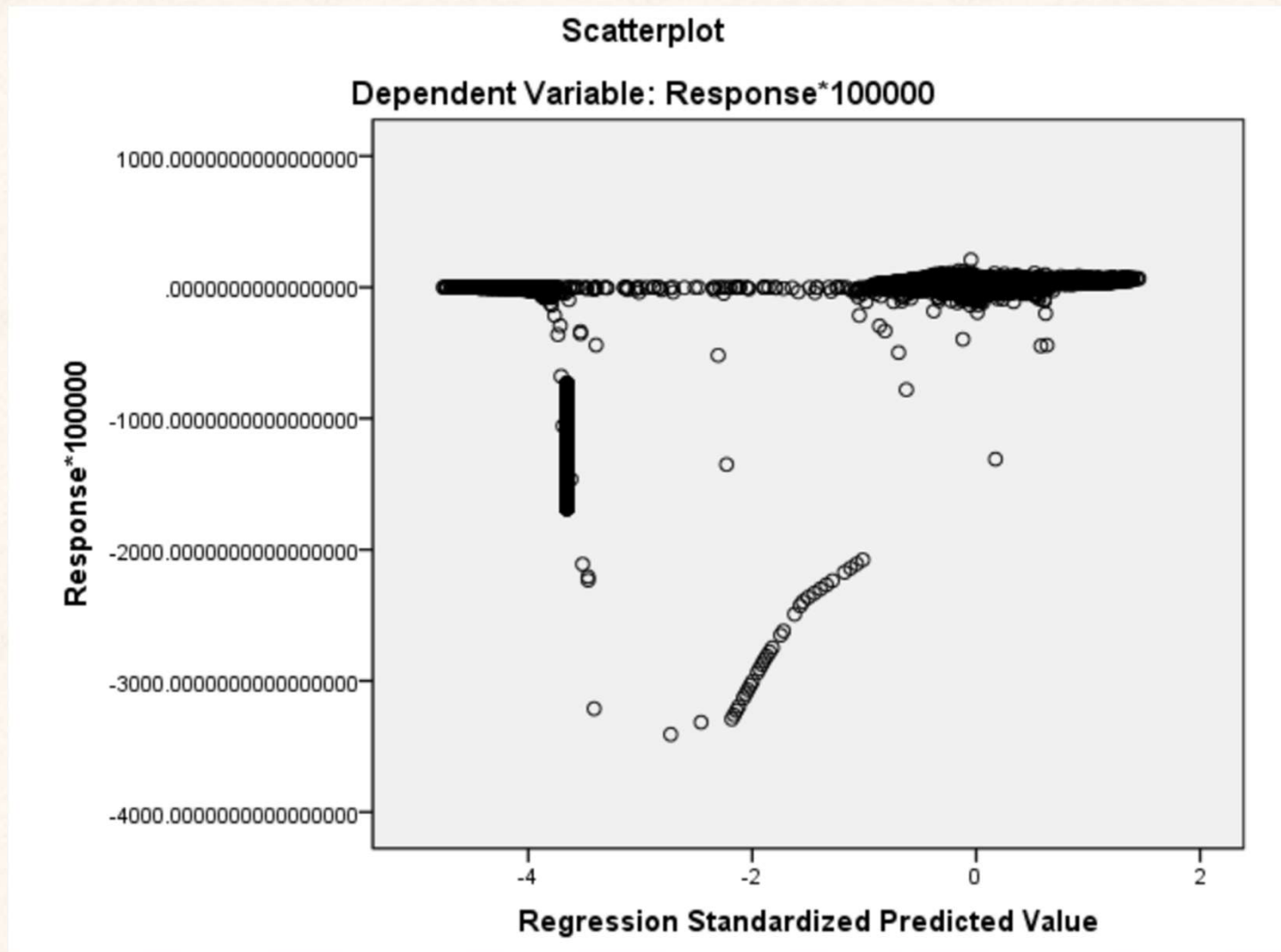
If condition index is greater than 30, strong multicollinearity may exist.

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-445.279083	147.8716431	9.110077238	95.46742114	48059
Residual	-3157.23145	444.3707275	.0000000000	164.2402986	48059
Std. Predicted Value	-4.760	1.453	.000	1.000	48059
Std. Residual	-19.223	2.706	.000	1.000	48059

Part C. Multiple Linear Regression



Part C. Multiple Linear Regression

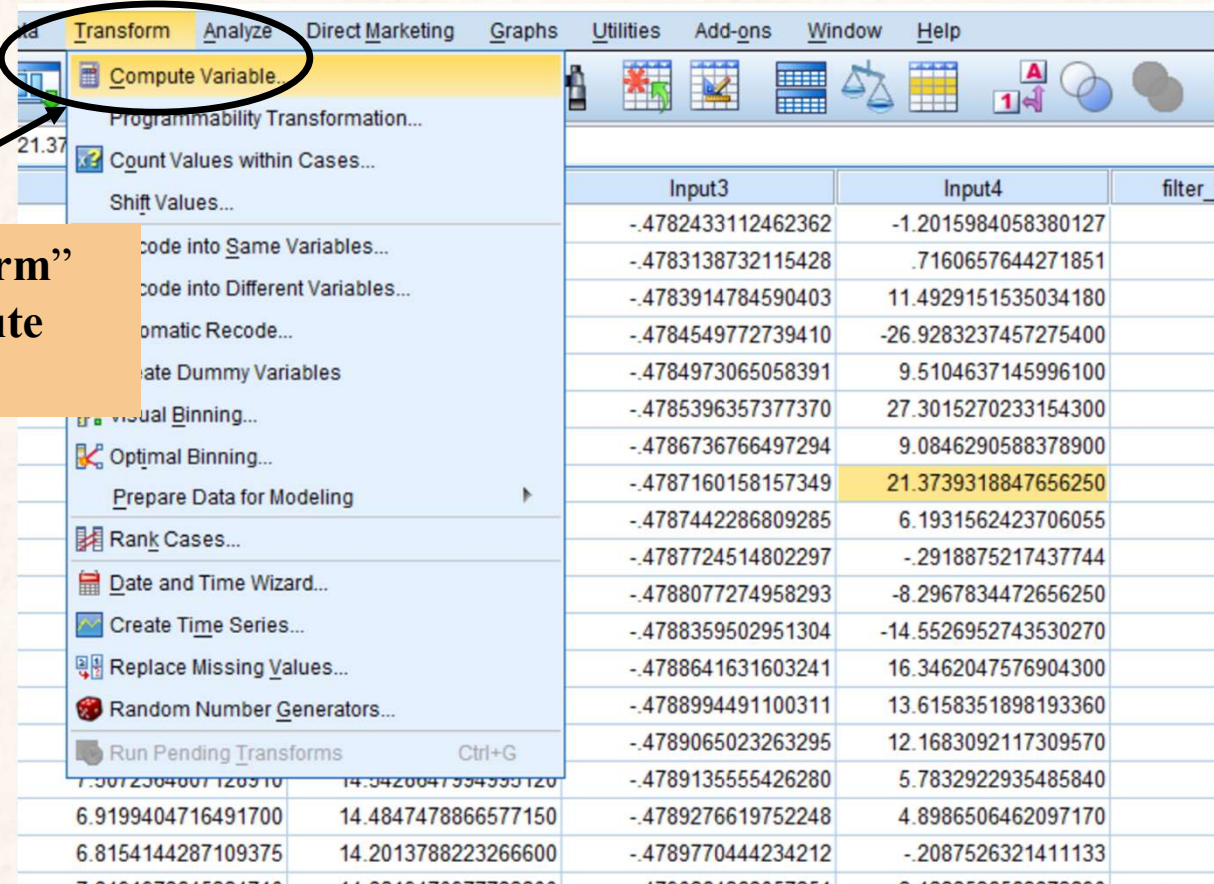


Part B. Simple Linear Regression

- Make prediction based on the current model for test set

1. Move to “Data_test”

2. Click the “Transform” tab and click “Compute Variable”



Part B. Simple Linear Regression

1. We put in **Pre12** as the **Target Variable**.

2. According to the results, Intercept is -332.688, the coefficient of Input1 is -17.944 and the coefficient of Input2 is 31.630. Put in the regression equation.

3. Click “OK”

The screenshot shows the 'Compute Variable' dialog box in SPSS. The 'Target Variable' field contains 'Pre12'. The 'Numeric Expression' field contains the regression equation:
$$-332.688 - 17.944 * \text{Input1} + 31.630 * \text{Input2}$$
. The left pane lists variables including 'Input 1 [Input1]', 'Input 2 [Input2]', 'Pre1', and 'Pre2'. The bottom of the dialog has buttons for 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'. Annotations with arrows point to the 'Target Variable' field, the 'Numeric Expression' field, and the 'OK' button.

Part B. Simple Linear Regression

1. We put in **Pre12** as the **Target Variable**.

2. According to the results, Intercept is -332.688, the coefficient of Input1 is -17.944 and the coefficient of Input2 is 31.630. Put in the regression equation.

3. Click **“OK”**

The screenshot shows the 'Compute Variable' dialog box in SPSS. The 'Target Variable' field contains 'Pre12'. The 'Numeric Expression' field contains the regression equation:
$$-332.688 - 17.944 * \text{Input1} + 31.630 * \text{Input2}$$
. The list of variables on the left includes 'Time', 'Response', 'Response*100000 [...]', 'Input 1 [Input1]', 'Input 2 [Input2]', 'Input 3 [Input3]', 'Input 4 [Input4]', 'Approximately 80% ...', 'Pre1', and 'Pre2'. The 'OK' button is circled at the bottom.

Part C. Multiple Linear Regression

-13.618131149291969

	Input2	Input3	Input4	filter_\$	Pre1	Pre2	Pre12	var
260	12.4994173049926760	-.4782433112462362	-1.2015984058380127	0	-8.1670205	-45.31	-58.41	
280	12.9199161529541020	-.4783138732115428	.7160657644271851	0	-14.0068715	-36.33	-40.70	
730	13.0335521697998050	-.4783914784590403	11.4929151535034180	0	-12.4848885	-33.90	-38.26	
320	12.7729568481445310	-.4784549772739410	-26.9283237457275400	0	-23.5529551	-39.47	-38.14	
680	12.5610017776489260	-.4784973065058391	9.5104637145996100	0	-21.2716125	-43.99	-46.57	
450	12.8111324310302730	-.4785396357377370	27.3015270233154300	0	-24.4258618	-38.65	-36.28	
830	12.5912876129150390	-.4786736766497294	9.0846290588378900	0	-31.5729083	-43.35	-37.83	
230	13.2391061782836910	-.4787160158157349	21.3739318847650	0	-29.52	-35.23	-35.23	
120	13.5111141204833980	-.4787442286809285	6.1931562423700	0	-23.71	-21.76	-21.76	
050	13.5423336029052730	-.4787724514802297	-.2918875217430	0	-23.04	-18.94	-18.94	
090	13.7622613906860350	-.4788077274958293	-8.2967834472650	0	-18.35	-26.88	-26.88	
840	13.5562973022460940	-.4788359502951304	-14.5526952743530	0	-22.74	-27.55	-27.55	
740	13.2583360671997070	-.4788641631603241	16.3462047576904300	0	-21.0153499	-29.11	-24.71	
760	14.1494235992431640	-.4788994491100311	13.6158351898193360	0	-4.3971697	-10.08	-9.07	
820	14.3734378814697270	-.4789065023263295	12.1683092117309570	0	11.0110210	-5.10	-13.62	
910	14.5428647994995120	-.4789135555426280	5.7832922935485840	0	9.8824750	-1.68	-7.41	
700	14.4847478866577150	-.4789276619752248	4.8986506462097170	0	-4.0746147	-2.92	1.29	
375	14.2013788223266600	-.4789770444234212	-.2087526321411133	0	-6.5586761	-8.97	-5.79	
710	14.8219470977783200	-.4790264368057251	-3.1222536563873290	0	18.0091873	4.27	-4.72	
840	15.3820934295654300	-.4790758192539215	-2.8295378684997560	0	21.4065540	16.23	10.44	
330	16.5982074737548830	-.4791322648525238	-6.3681745529174805	0	37.1308066	42.19	37.03	

le View

Exercise

- 1. Split the dataset to training set (60%) and test set (40%).
- 2. Based on Ex.1, do **Simple Linear Regression** using **Input3** as X, **Response*100000** as Y. Then make predictions for the test set.
- 3. Do **Multiple Linear Regression** using **Input3** and **Input 4** as X, **Response*100000** as Y. Then make predictions for the test set.

End of Lab 3



Please let me check your work,
then you can leave.

See you in the next Lab.