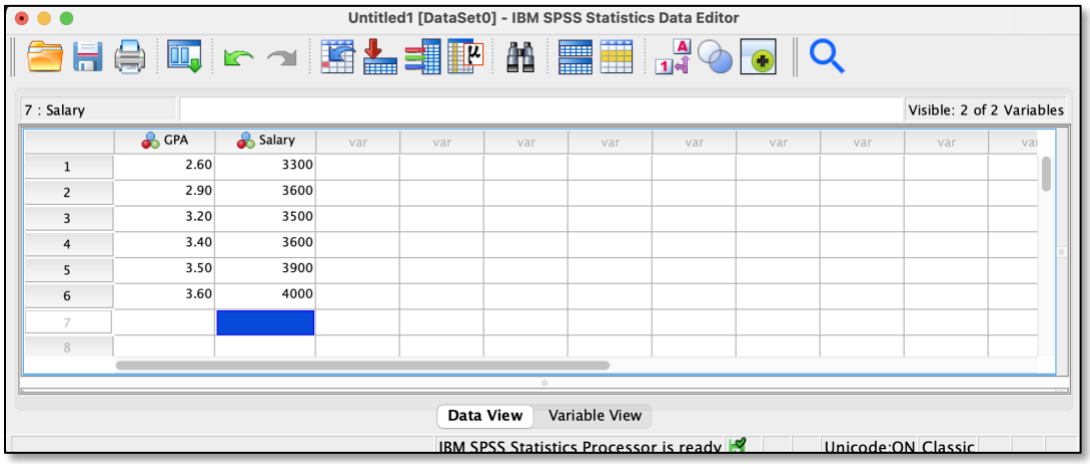
Introduction to SPSS for Linear Regression

IBM SPSS is a powerful statistical tool that offers wide range of tools to perform analyses such as Regression, Classification, Significance tests etc. It is a closed source (IBM proprietary) software which makes it less used as compared to R or Python. However, there are a good number of tutorials on YouTube that covers various concepts that can be applied with IBM SPSS. In this document, we aim to cover the basic concepts required to perform and evaluate Linear Regression in IBM SPSS.

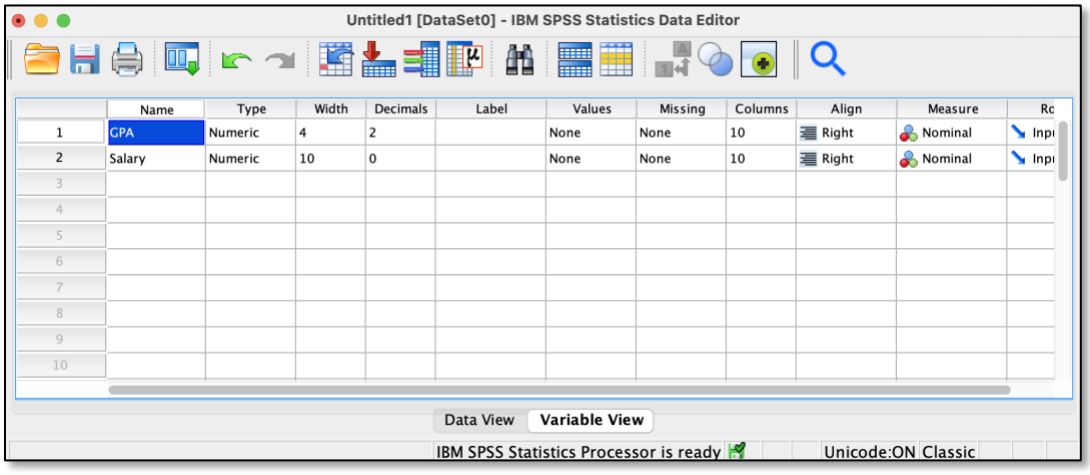
1. SPSS Windows

As mentioned earlier, IBM SPSS looks like an extension of Microsoft Excel with two editors – Data Editor and Statistics Viewer. In Data Editor, there are two views – Data view and the Variable view. Data view is where the entire data is listed in rows and columns format and the Variable view shows information on the type of variable (column) present in the data.
All the analysis, when performed, will be displayed on the Statistics Viewer.

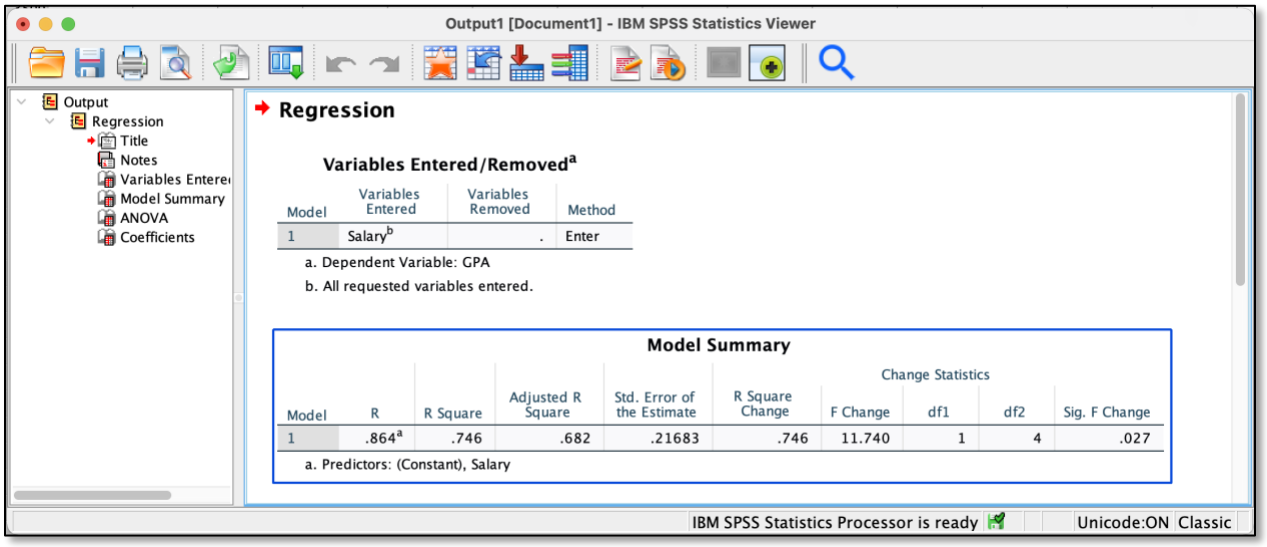
Data Editor – Data View



Data Editor – Variable View

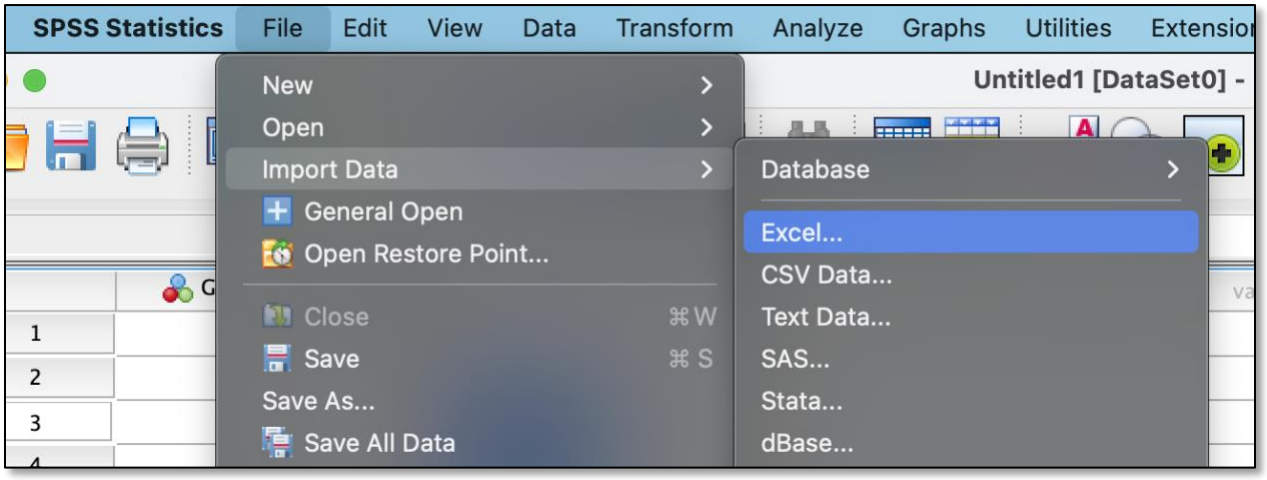


Statistics Viewer



2. Data Loading

To import data into SPSS, click on File -> Import Data -> Excel/CSV. The data can be copied and pasted from Excel sheets to SPSS as well. But care should be taken to ensure proper the types of variables.

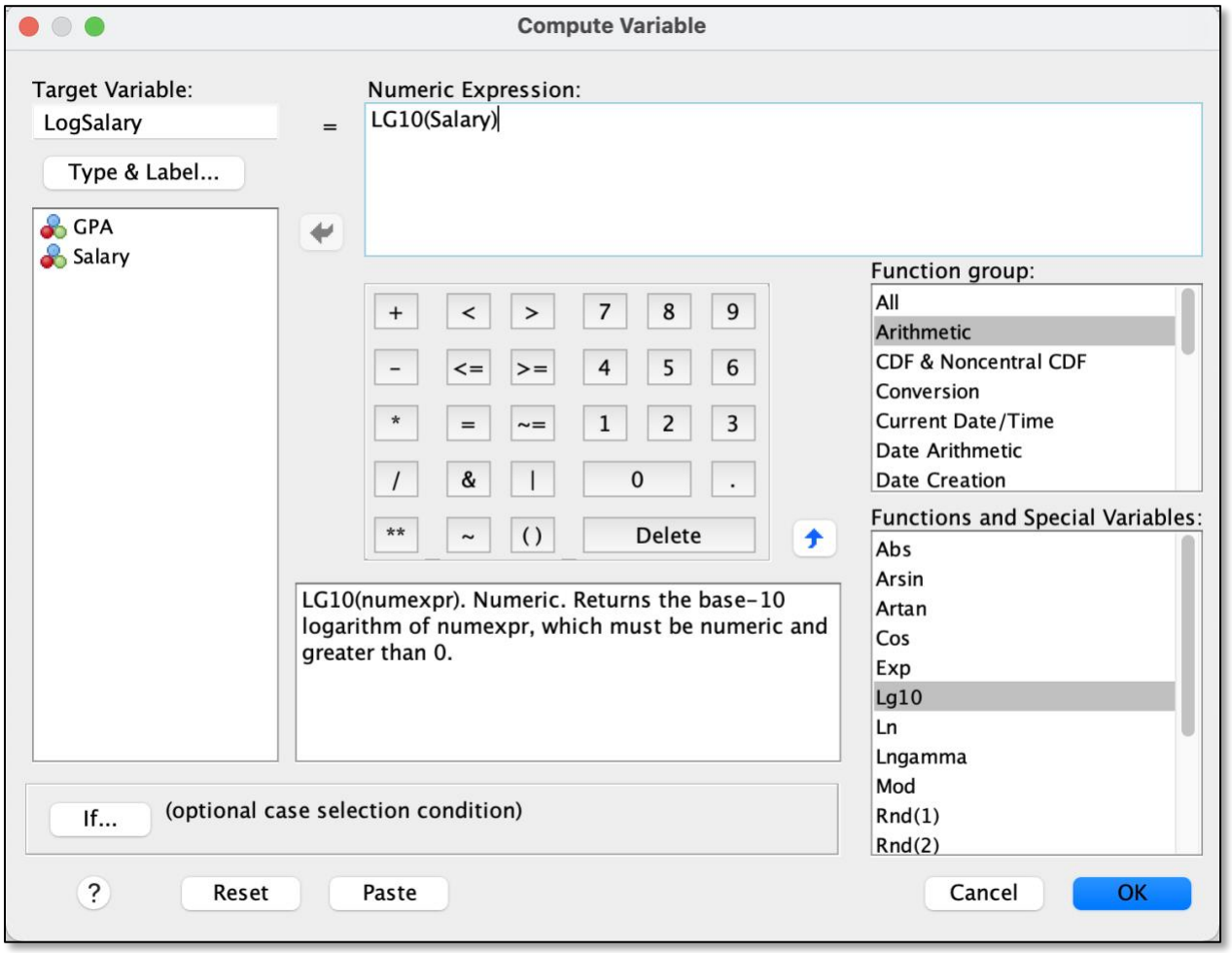


3. Data Transformation

At times, it becomes necessity to apply transformations such as squaring, log transformations etc. to the original data to infer linear relationships between variables. In such times, transformation can be applied by using the following steps –

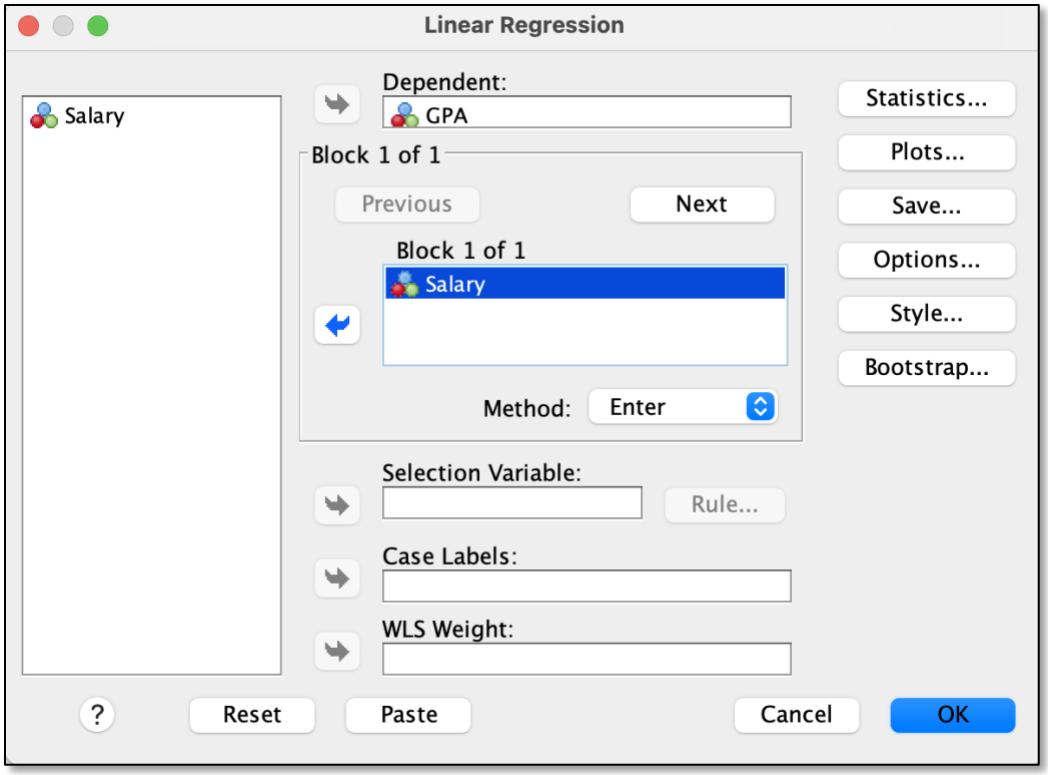
a. Select the variable to be transformed
b. Click on Transform -> Compute Variable

c. Enter the name of the Target variable (Transformed variable) and the formula to transform into. Note – Most of the arithmetic functions can be found under Arithmetic group

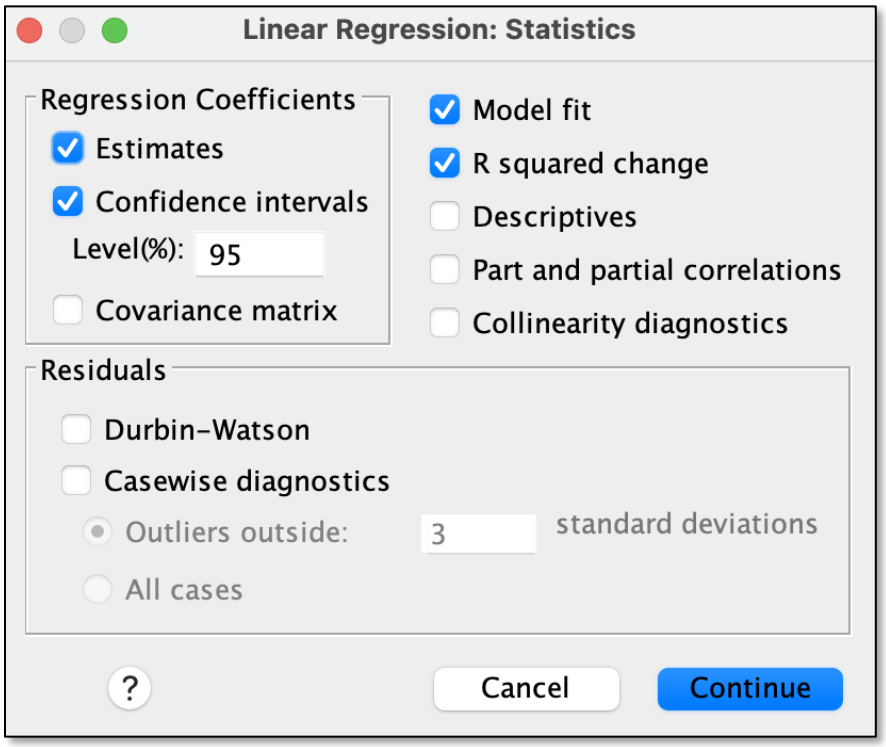


4. Linear Regression

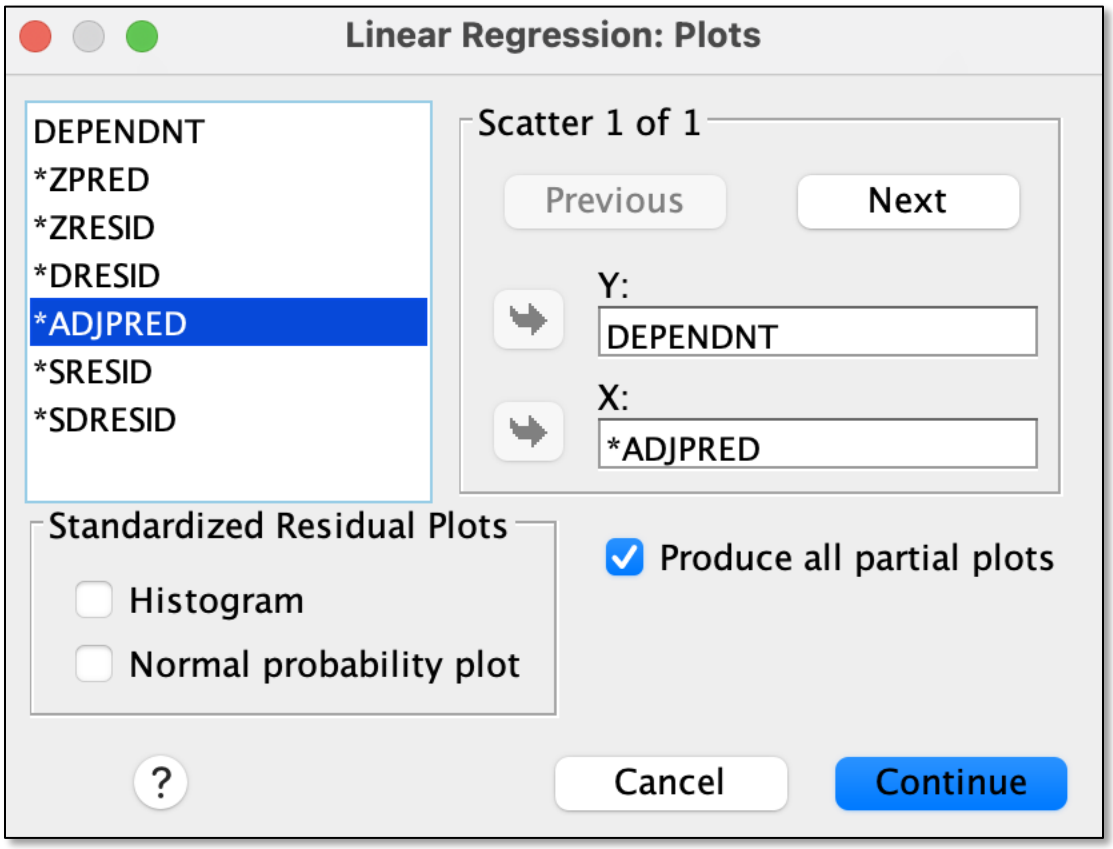
To perform regression, select the data and click on Analyze -> Regression -> Linear. In the Linear Regression window, enter the dependent variable (Y) and the list of independent variables (X).
The “Method” dropdown allows to configure the type of regression with respect to feature selection such as Stepwise, Backward, Forward to be considered.



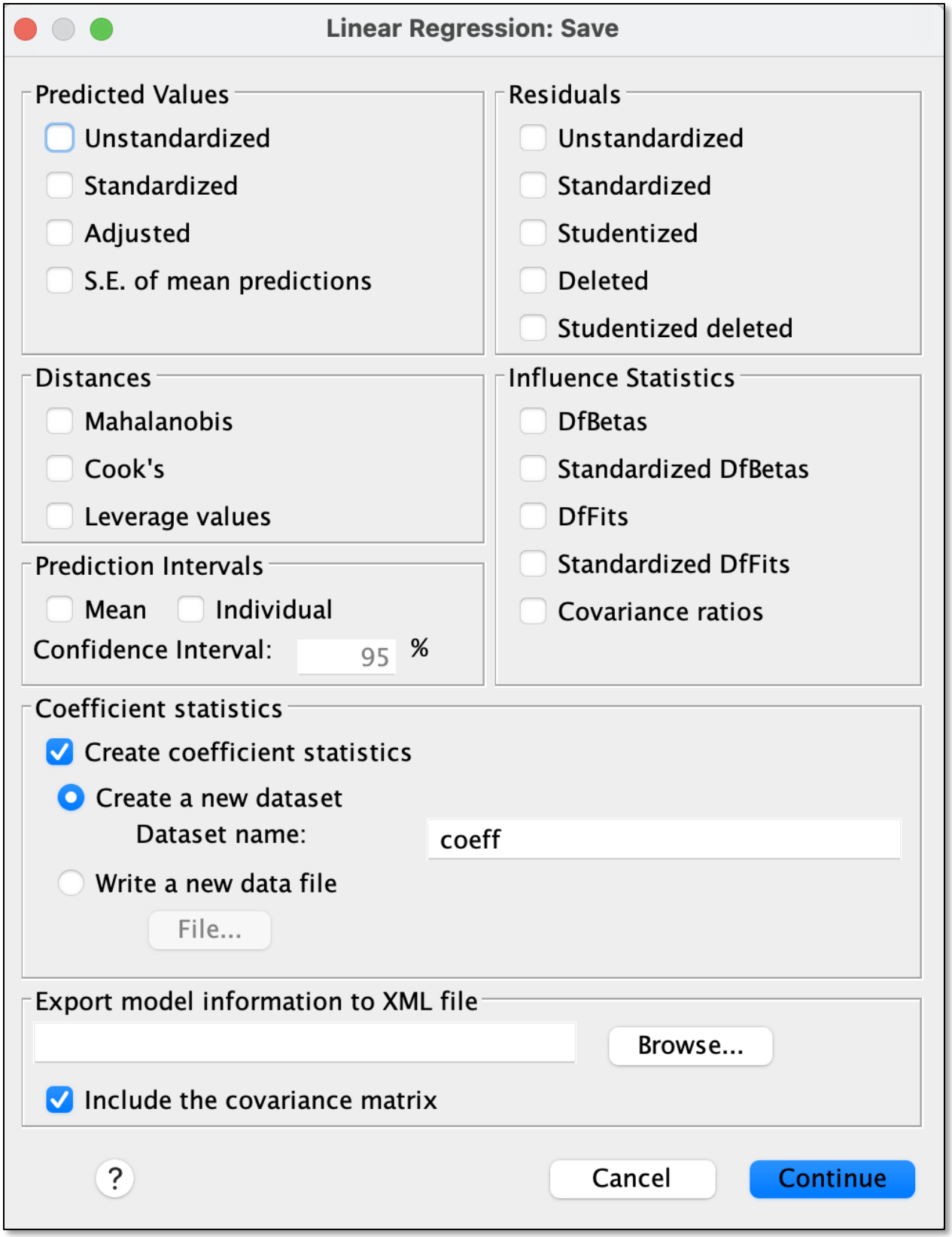
Under Statistics, the list of functions that needs to be analyzed can be specified.



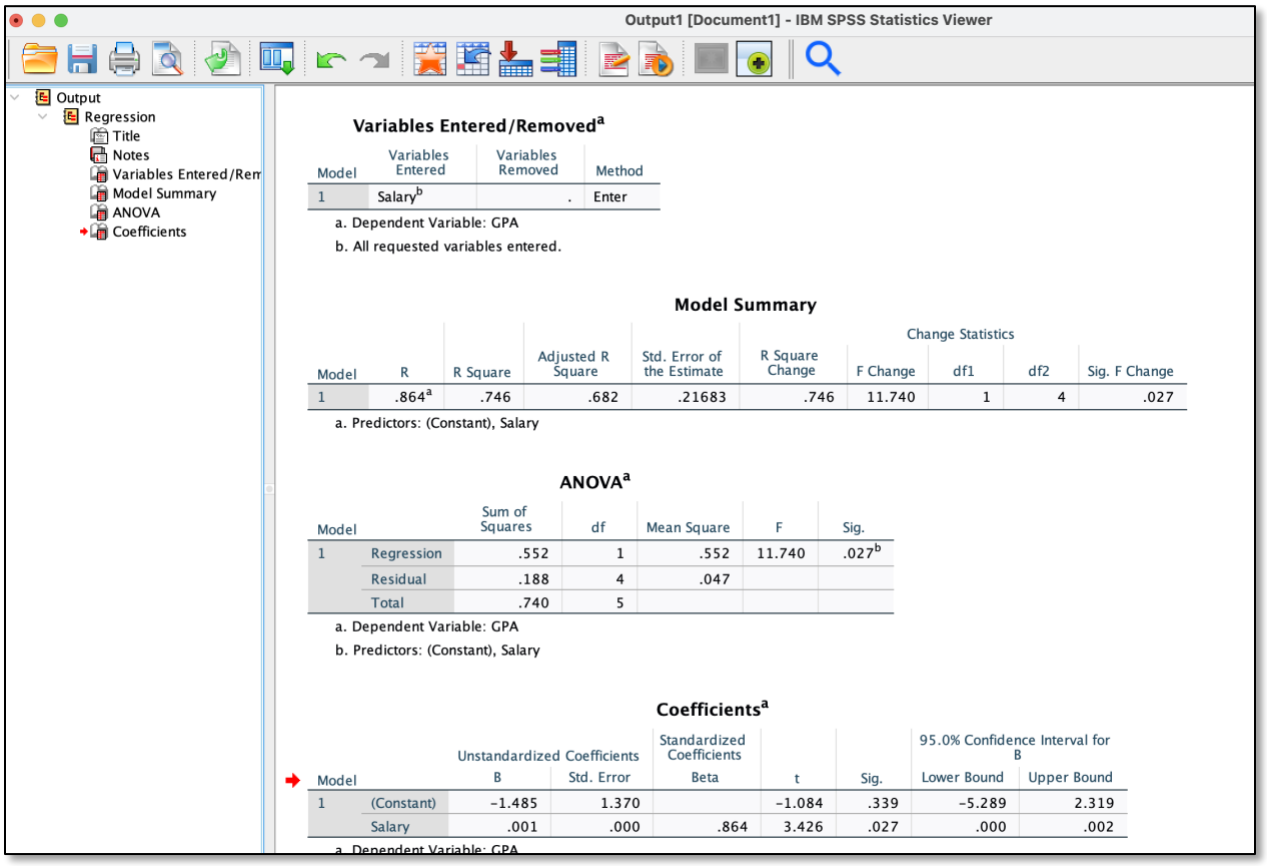
Under plots, the list of plots and the variables to plot can be specified. More plots can be added by clicking on the “Next” button.



Under the Save options, specify the list of metrics and analyses that needs to be displayed. Note that, some of the values can also be stored into a dataset/XML file in this window.

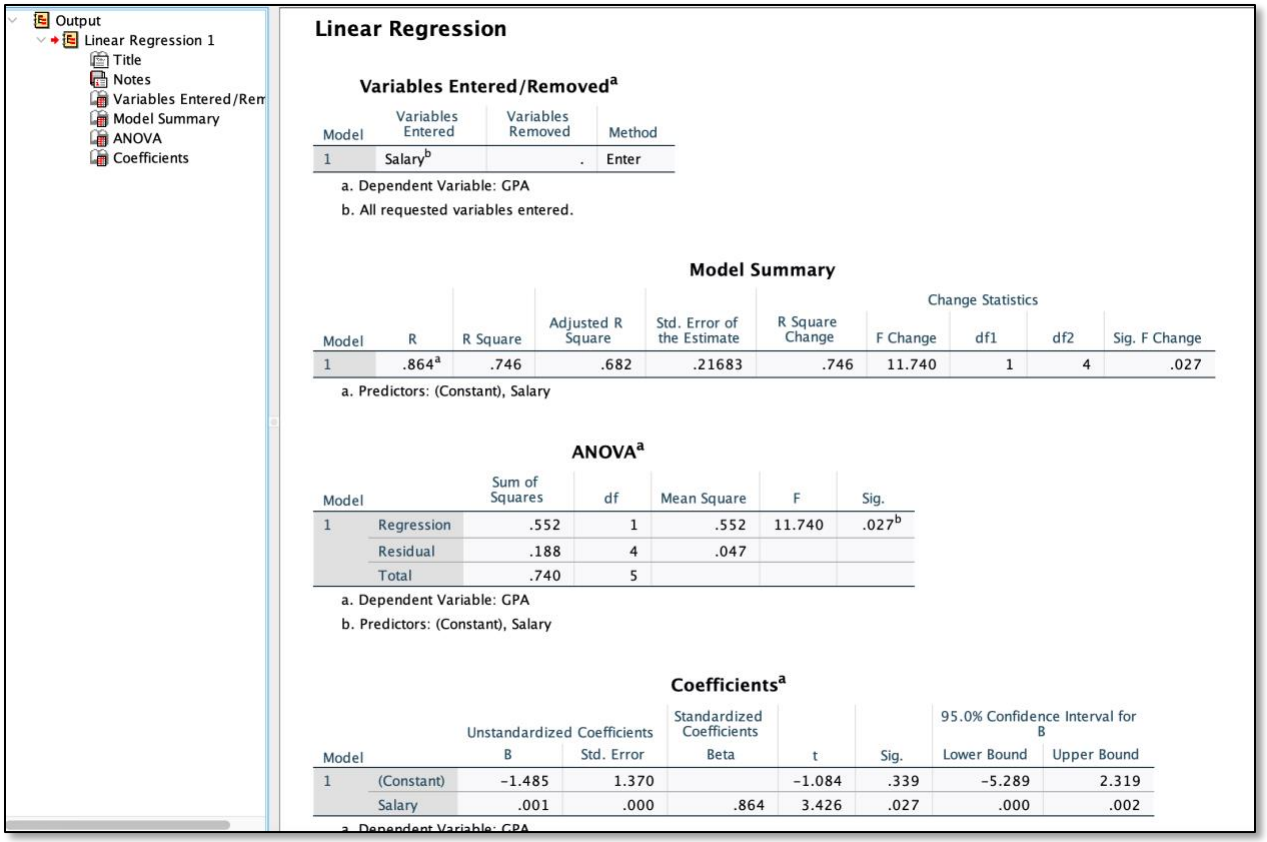

Linear Regression: Save
Predicted Values
Unstandardized
Standardized
Adjusted
S.E. of mean predictions
Residuals
Unstandardized
Standardized
Studentized
Deleted
Studentized deleted
Distances
Mahalanobis
Cook's
Leverage values
Influence Statistics
DfBetas
Standardized DfBetas
DfFits
Standardized DfFits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

On clicking OK, the analyses will get populated on the Statistics Viewer window. On the left-hand side, the navigation view can be seen of all the analyses performed and on the right side, the analyses can be seen.



5. Evaluating the model

Once the analysis is performed, various results are listed in the Statistics Viewer window. This can be used to evaluate the performance of the model by looking at things like Significance, R Squared etc. Another advantage of SPSS is multiple models results can be listed one below another in the Statistics Viewer window that allows to quickly compare between the models. Care should be taken to note which is the model that one is currently looking at, which can be observed by looking at the Navigation Menu. The models could also be renamed as per the user's choice.

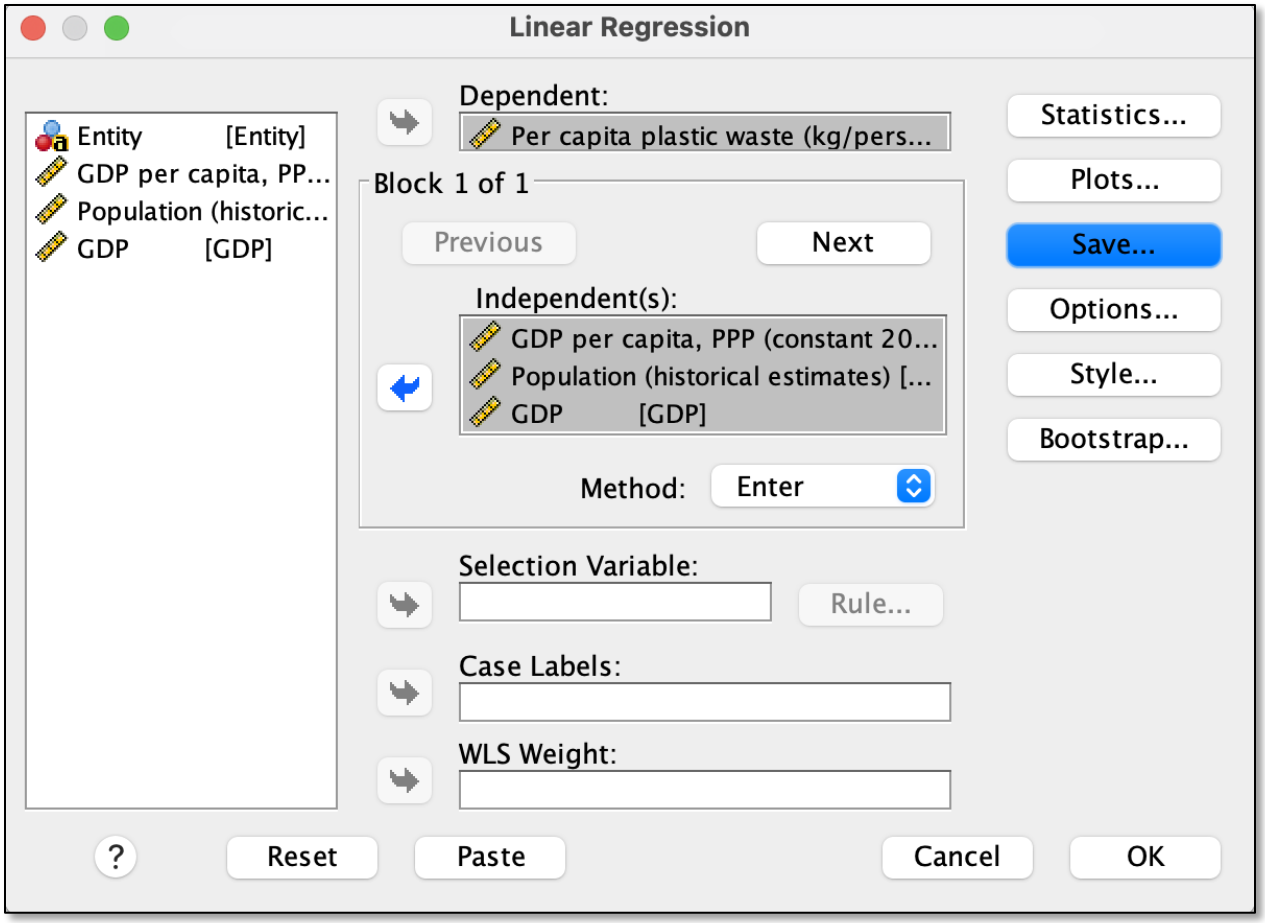


6. Making predictions

To make predictions on the new dataset, the following steps are to be performed –

a. Build the model using the training dataset. If predictions on training dataset is required, then please make sure to check the Unstandardized Predicted values under Save option in the Linear Regression window.

Entity	GDPpercapitaPPPconstant2017international$	Populationhistoricalestimates	GDP	Percapitaplasticwastekgpersonday
Germany	46929.99	80827001.00	3.79E+12	.48500
Netherla	52032.99	16682927.00	868063000000	.42400
Sri Lank	9126.87	20261738.00	184926000000	.35700
United S	54315.91	309011469.00	1.68E+13	.33500
Guatemal	7335.99	14630420.00	107329000000	.28000
Spain	37319.48	46931011.00	1.75E+12	.27700
Portugal	31798.15	10596055.00	336935000000	.26500
South Af	12452.34	51216967.00	637771000000	.24000
United K	42089.01	63459801.00	2.67E+12	.21500
Turkey	20027.67	72326992.00	1.45E+12	.21200
Greece	33753.62	10887640.00	367497000000	.20000
Malaysia	20536.37	28208028.00	579291000000	.19800
France	42147.67	62879535.00	2.65E+12	.19200
Argentin	23521.27	40895751.00	961920000000	.18300
Egypt	10340.07	82761244.00	855757000000	.17800
Japan	37586.17	128542349.00	4.83E+12	.17100



b. Note down the unstandardized coefficients

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	.181	.047		3.842	<.001
	GDP per capita, PPP (constant 2017 international $)	1.895E-6	.000	.320	1.266	.219
	Population (historical estimates)	-6.091E-10	.000	-.427	-.918	.368
	GDP	1.397E-14	.000	.494	.933	.361

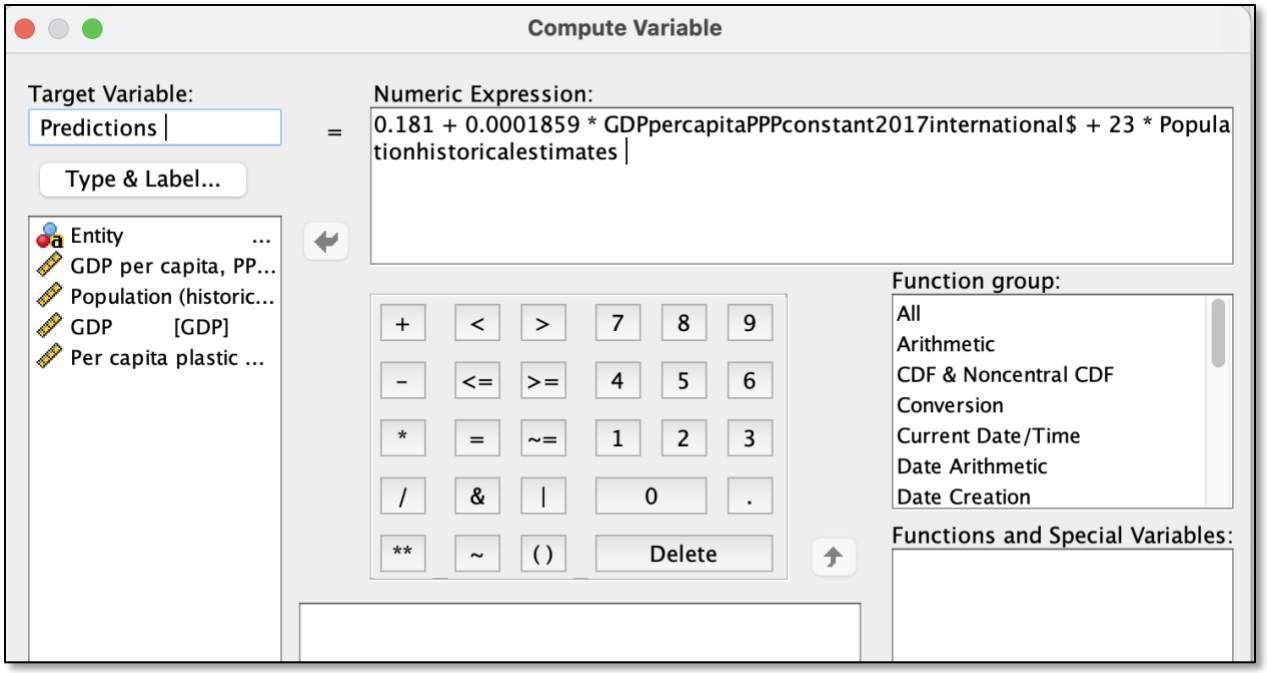
a. Dependent Variable: Per capita plastic waste (kg/person/day)

c. Import the new dataset (test dataset)

Entity	GDPpercapitaPPPconstant2017international$	Populationhistoricalestimates	GDP	Percapitaplasticwastekgpersonday	var	var
Nigeria	4932.33	158503203.00	781791000000.00	.10300		
Vietnam	5089.41	87967655.00	447704000000.00	.10300		
Ukraine	11778.31	45792086.00	539354000000.00	.10300		
Sudan	3089.57	34545014.00	106729000000.00	.10300		
Iraq	8748.51	29741977.00	260198000000.00	.10300		
Cote d'I	3660.90	20532944.00	75169133689.00	.10300		
Senegal	2797.11	12678143.00	35462162599.00	.10300		
Poland	23996.14	38329784.00	919767000000.00	.09700		
Canada	44861.52	34147566.00	1531910000000.00	.09300		
Mexico	17790.01	114092961.00	2029720000000.00	.08700		
Belgium	47971.63	10938735.00	524749000000.00	.08000		
Philippi	5918.37	93966784.00	556130000000.00	.07500		
Myanmar	3129.92	50600827.00	158377000000.00	.07500		
Morocco	6281.46	32343384.00	203164000000.00	.07300		
Cambodia	2716.70	14312205.00	38881963567.00	.06600		
Angola	7692.43	23356247.00	179666000000.00	.06200		

d. Create the predictions using the linear regression formula and the coefficients

To do this, click on Transform -> Compute Variable -> Enter the expression based on the unstandardized co-efficients and the model



Entity	GDPperCapita	Population	GDP	Percapitaplasticwastekgperson day	Predictions
Nigeria	4932.33	158503203.00	781791000000.00	.10300	.20
Vietnam	5089.41	87967655.00	447704000000.00	.10300	.20
Ukraine	11778.31	45792086.00	539354000000.00	.10300	.21
Sudan	3089.57	34545014.00	106729000000.00	.10300	.19
Iraq	8748.51	29741977.00	260198000000.00	.10300	.20
Cote d'I	3660.90	20532944.00	75169133689.00	.10300	.19
Senegal	2797.11	12678143.00	35462162599.00	.10300	.19
Poland	23996.14	38329784.00	919767000000.00	.09700	.24
Canada	44861.52	34147566.00	1531910000000.00	.09300	.29
Mexico	17790.01	114092961.00	2029720000000.00	.08700	.24
Belgium	47971.63	10938735.00	524749000000.00	.08000	.28
Philippi	5918.37	93966784.00	556130000000.00	.07500	.20
Myanmar	3129.92	50600827.00	158377000000.00	.07500	.19
Morocco	6281.46	32343384.00	203164000000.00	.07300	.20
Cambodia	2716.70	14312205.00	38881963567.00	.06600	.19
Angola	7693.43	33256247.00	170666000000.00	.06300	.20

This would create the predictions on the new column. These predictions can be used to evaluate the model's performance and help in selecting the best model.