Tutorial overview

In this tutorial, we show how to perform a regression analysis.

Our dataset consists in engine cars description. We want to predict "mpg" consumption from cars characteristics such as weight, horsepower, ...

In this tutorial, we use the following components:

ТАВ	Operator (Component)	Comment		
Data visualization	View dataset	View the dataset in a grid		
Feature selection	Define status	Define attributes status		
Regression	Multiple linear regression	Perform the regression analysis		

Loading the dataset

- Open the existing stream diagram
 - 1 Choose *File/Open...* in the TANAGRA main menu.
 - 2 Get « autompg.bdm » which is in the « Dataset » sub-directory.
- > Add a component for data visualization

1 – Add **View dataset** to the diagram. Click on DATA VISUALIZATION components tab, select **View dataset** component. Drag this component over the diagram and drop on the "Dataset" node.

2 – Select the "View dataset" component in the diagram (if it is not already selected) and right click to show the popup menu. Choose the *View* command, dataset will be visible in the right frame.

Tanagra Begre	ssion	<u> </u>	anlinder	dign logoment	horgonomor	waight	logge lovet i	
			cylinder	aisplacement	norsepower	s 504	acceleraci	mpg
Dataset (autompg.txt)				350	165	3603	11 5	
			8	318	150	3436	11.5	
		4	8	304	150	3433	12	16
		5	8	302	140	3449	10.5	
		6	8	429	198	4341	10	
		7	8	454	220	4354	9	
		8	8	440 455 390	215 225 190	4312 4425 3850	8.5 10 8.5	14 14 15
		9	8					
		10	8					
			8	383	170	3563	10	15
1			8	340	160	3609	8	14
			8	400	150	3761	9.5	15
		14	8	455	225	3086	10	
		15	4	113	95	2372	15	24
		16	6	198	95	2833	15.5	
			6	199	97	2774	15.5	18
			Com	ponents				
Data visualization	Descriptive stats		Instan	ce selection	Feature construction			
Feature selection	Regression		Facto	rial analysis	Clustering			
Spv learning	Meta-spy learning Spy le			ning assesment	Associa			
Correlation scatternlot	Wiew dataset	3,0520	0.0 532					
To one of the source place in the second of								

There are 392 cars in the dataset, described with 6 continuous attributes: number of cylinder, displacement, horsepower, weight, acceleration and consumption (miles per gallon).

Define status of attributes for regression

1 – Click on the "Dataset" component in the diagram and add a **Define Status** component with the shortcut in the tool bar. A dialog box appears automatically, you may define the status of each attribute.

Tanagra Begres	sion		aulinder	dienlagene	horeenover	weight	aggalarat
Hangid Hegission Hangid Hegission I I I View dataset 1 View dataset 1			eyiinder 8	displaceme 307 350	130 165	werght 2504	accelerat
						3693	11 5
			8				11
Define status 1		4	8	304	150	3433	12
		5	8	302	140	3449	10 5
		6	8	429	198	4341	10
		7	8	454	220	4354	
		8	8	440	215	4312	8.5
		9	8	455	225	442.5	10
		10	8	390	190	3850	8.5
		11	8	383	170	3563	10
		12	8	340	160	3609	8
		13	8	400	150	3761	9.5
		14	8	455	225	3086	10
		15	4	113	95	2372	15
		16	6	198	95	2833	15.5
			Components				
Data visualization	Descriptive stats		Instance selec	tion Fe	ature construc	tion	
Feature selection	Regression		Factorial anal	ysis	Clustering		
Spy learning	Meta-spv learning		Sov learning asse	sment	Association		
	TOTA //	-+					
correlation scatterplot		el					

2 – Ensure that "Input" tab is selected, choose the five first attributes of the list, and click on the arrow button to add them to the "Input" list. These attributes are "exogenous" attributes of the analysis.

Define attributes status Parameters Report	
Attributes :	Target Input Illustrative
C cylinder C displacement C horsepower C weight C acceleration C mpg	cylinder displacement horsepower weight acceleration Add selected attributes
E. E.	Clear selection
	OK Cancel Help

3 – Activate the "Target" tab. Select "mpg" attribute and add this one to the target list. This attribute is the "endogenous" attribute; we want to predict its values from exogenous ones.

Define attributes status Parameters Report			1
Attributes : C cylinder C displacement C horsepower C weight C acceleration	mpg	Input Illust	rative
E. E.	Clea	ar selection	
	ок	Cancel	Help

4 – Click on the OK button to validate this selection and close this dialog box.

5 – Double-click this **Define Status** component, you can verify on the right side of the window your attribute selection.

Define status 1 Results Attribute Target Input Illustrative cylinder Period Colspan="2">Define status 1 Define status 1 Norsepower Period Components Deta visualization Descriptive stats Instance selection Feature selection Regression Factorial analysis Clustering Spy learning Spy learning assesment Association Association	Tanagra - [Define status <u>P</u> File <u>D</u> iagram Component	1] indowHelp			_			_ D × _ B ×	
Define status 1 Define status 1 Parameters Target : 1 Input : 5 Illustrative : 0 Results Attribute Target Input Illustrative cylinder - yes - displacement - yes - mpg yes - execution time : 0 ms. Components Data visualization Descriptive stats Instance selection Feature construction Feature selection Regression Data visualization Descriptive stats Instance selection Feature construction Feature selection Regression Components Data visualization Meta-spy learning Spy learning Meta-spy learning Spy learning View multiple scatterplot View multiple scatterplot	D 📽 🖬 🔤								
Data visualization Descriptive stats Instance selection Feature construction Data visualization Descriptive stats Instance selection Feature construction Feature selection Regression Factorial analysis Clustering Spy learning Meta-spy learning Spy learning Spy learning Weiw dataset Export dataset Export dataset Export dataset	Tanagra Regress	Define status 1							
Results Attribute Target Input Illustrative cylinder - cylinder - yes - - displacement - yes - - horsepower - yes - - weight - yes - - mpg yes - - - Execution time : 0 ms. - - - Components - - - - Spv learning Meta-spv learning Sp	View dataset 1	.)	Parameters Target : 1 Input : 5 Illustrative : 0						
Attribute Target Input Illustrative cylinder cylinder - yes cylinder - yes displacement - yes horsepower - yes weight - yes acceleration - yes mpg yes - Execution time : 0 ms. - Execution time : 0 ms. - Execution time : 0 ms. - Components - Data visualization Descriptive stats Instance selection Feature selection Regression Factorial analysis Spv learning Meta-spv learning Spv learning assesment Association Spv learning View dataset Export dataset - - View multiple scatterplot View multiple scatterplot							Results		
cylinder - yes - displacement - yes - horsepower - yes - weight - yes - acceleration - yes - mpg yes - - Execution time : 0 ms. - - Components - - Data visualization Descriptive stats Instance selection Feature selection Regression Factorial analysis Clustering Spv learning Meta-spv learning Spv learning assesment <td< td=""><td></td><td></td><td>Attribute</td><td>Target</td><td>Input</td><td>Illustrative</td><td></td><td></td></td<>			Attribute	Target	Input	Illustrative			
displacement - yes horsepower yes weight yes acceleration yes mpg yes mpg yes Execution time : 0 ms. Execution time : 0 ms. Components Execution Data visualization Descriptive stats Instance selection Feature construction Feature selection Regression Spv learning Meta-spv learning Spv learning Meta-spv learning Spv learning View dataset Export dataset i: View multiple scatterplot i: View multiple scatterplot i: View multiple scatterplot			cylinder	-	yes	-			
horsepower - yes - weight - yes - acceleration - yes - mpg yes Execution time : 0 ms. Constructed of or for for on an or Execution time : 0 ms. Constructed of or for on an or Execution time : 0 ms. Constructed of or for on an or Execution time : 0 ms. Constructed of or for on an or Execution time : 0 ms. Constructed of or for on an or Execution time : 0 ms. Components Execution time : 0 ms. Components Components Components Components Components Components Clustering Spv learning Meta-spv learning Spv learning assesment Association			displacement	-	yes	-			
weight - yes - acceleration - yes - mpg yes - - mpg yes - - Execution time : 0 ms. - - Execution time : 0 ms. - - Data visualization Descriptive stats Instance selection Feature construction Feature selection Regression Factorial analysis Clustering Spv learning Meta-spv learning Spv learning assesment Association Correlation scatterplot Istemplot Istemplot Istemplot Export dataset Istemplot Istemplot Istemplot			horsepower	-	yes	-			
acceleration - yes - mpg yes - mpg yes - Execution time : 0 ms. Execution time : 0 ms. Components Data visualization Descriptive stats Instance selection Feature selection Regression Factorial analysis Clustering Spv learning Meta-spv learning Spv learning Spv learning View dataset Export dataset View multiple scatterplot			weight	-	yes	.			
mpg yes Execution time : 0 ms. Execution time : 0 ms. Execution time : 0 ms. Components Data visualization Descriptive stats Instance selection Feature selection Regression Factorial analysis Clustering Spv learning Meta-spv learning Spv learning			acceleration	-	yes	-			
Execution time : 0 ms. Components Data visualization Descriptive stats Instance selection Feature construction Feature selection Regression Spv learning Meta-spv learning Spv learning Meta-spv learning Spv learning View dataset Export dataset View multiple scatterplot			mpg	yes Da	-	-			
Components Data visualization Descriptive stats Instance selection Feature construction Feature selection Regression Factorial analysis Clustering Spv learning Meta-spv learning Spv learning assesment Association Correlation scatterplot Silview dataset Export dataset Silview multiple scatterplot			Execution time	e:0ms.	0.00.4	•			
Data visualization Descriptive stats Instance selection Feature construction Feature selection Regression Factorial analysis Clustering Spv learning Meta-spv learning Spv learning assesment Association Correlation scatterplot Sive with white the scatterplot It was a set				Compo	onents				
Feature selection Regression Factorial analysis Clustering Spv learning Meta-spv learning Spv learning assesment Association Correlation scatterplot Intervention scatterplot Intervention scatterplot Intervention scatterplot	Data visualization	Descriptive st	ats	Instance selection			Feature construction		
Spy learning Meta-spy learning Spy learning assesment Association Correlation scatterplot Interview dataset Interview dataset Export dataset Interview multiple scatterplot	Feature selection	Regression		Factorial analysis			Clustering		
Correlation scatterplot View dataset Export dataset View multiple scatterplot	Spv learning	Meta-spv learr	ning Sp	Spv learning assesment		esment	Association		
	Correlation scatterplot Export dataset Scatterplot	🔛 View dataset 🔛 View multiple	scatterplot						

Perform the regression analysis

1 – Add **Multiple linear regression component** (REGRESSION TAB) to the stream diagram, after the node «Define status 1».

2 – There is no parameter to specify in this analysis, select *View* in the popup menu of "Multiple linear regression 1" to run the analysis. Results appear in the right frame.



Coefficient of determination is rather good (0.70), the most significant attributes seem to be "horsepower" and "weight". I think nevertheless that there is a very strong colinearity between the exogenous attributes, it would be necessary to study more in detail this regression.