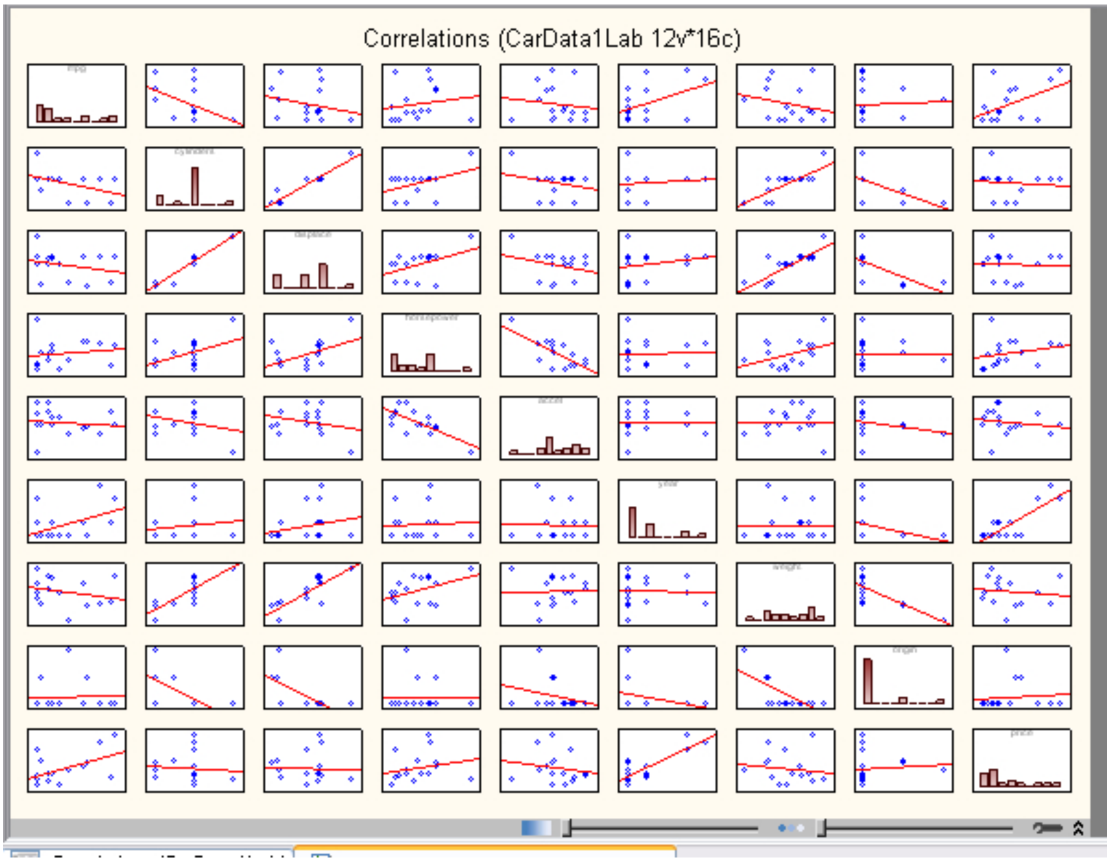
REGRESSION ANALYSIS

First, the dependence of the output (target) characteristic of the object on one of the input variables x_i (pair regression) is built, and then the linear model on all input variables x_1–x_n using the procedures of regression.

PAIR REGRESSION

Pairwise regression analysis is performed for the first sample - a car with mpg in the range of 20.1-22.5. A set of two-dimensional scattering diagrams is constructed. as the input variable I selected variable *cylinders* , since it has the greatest impact on the target variable – price.

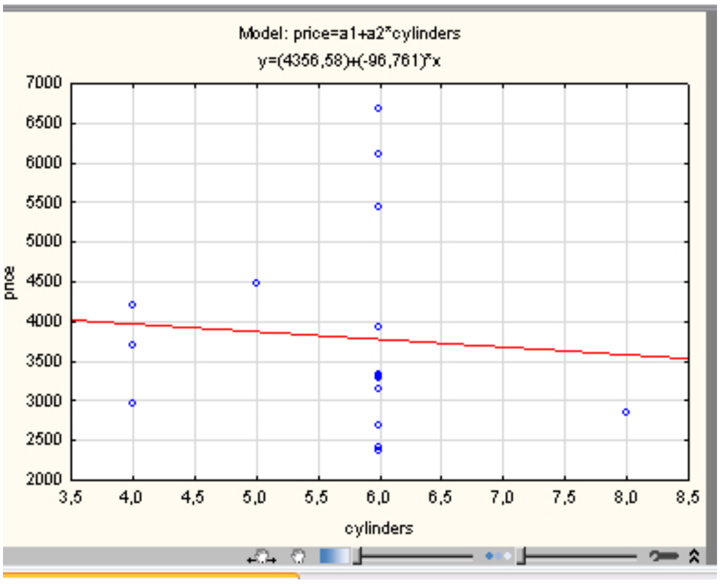


Picture 1. The set of two-dimensional scattering diagrams

Next, the following models were constructed using the least squares method:

Linear model

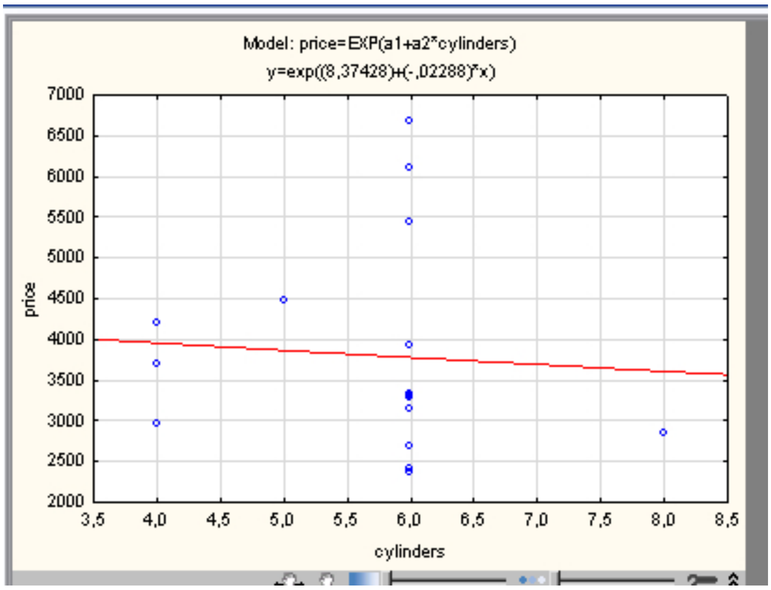
Model: price=a1+a2*cylinders and R^2=0,005826



Picture 2.. Linear regression model

Exponential model

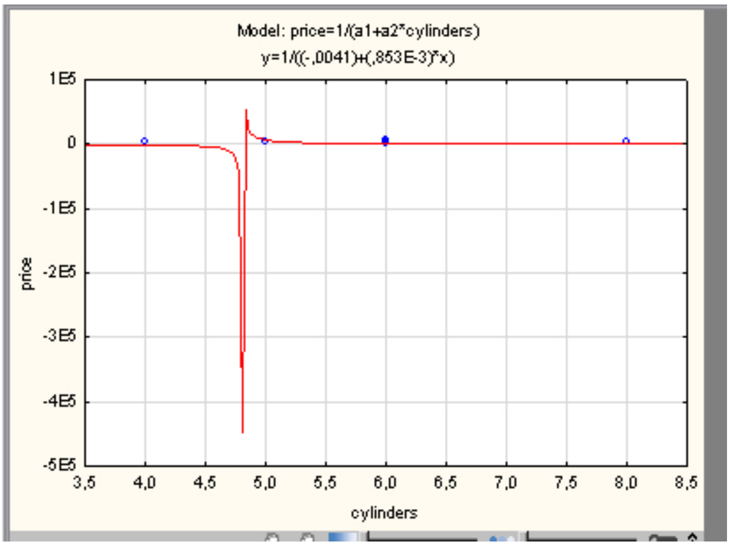
Model: price=EXP(a1+a2*cylinders) and R^2=0,005252



Picture 3. Exponential regression model

Inverse model

Model: price=1/((a1+a2*cylinders) and R^2=0



Picture 4. Inverse regression model

Coefficient of determination R^2 shows the discrepancies between the observed and estimated values of the output variable, the closer it is to unity, the more consistent the model with the data. The best is a linear model.

MULTIVARIATE REGRESSION

In this part I analyzed multivariate regression . Variables that would be part of the model were planned

	Regression Summary for Dependent Variable: price (CarData1Lab) R= ,96775587 R?= ,93655143 Adjusted R?= ,86403878 F(8,7)=12,916 p<,00148 Std.Error of estimate: 474,19					
N=16	b*	Std.Err. of b*	b	Std.Err. of b	t(7)	p-value
Intercept			-96987,4	12140,04	-7,98905	0,000092
mpg	0,24992	0,206438	412,2	340,52	1,21064	0,265320
cylinders	0,94437	0,446550	1197,1	566,07	2,11482	0,072272
displace	-1,13161	0,385215	-29,9	10,19	-2,93760	0,021788
horsepower	0,14226	0,197262	12,6	17,48	0,72119	0,494167
accel	0,09244	0,224734	86,1	209,41	0,41132	0,693136
year	1,03681	0,144605	1017,6	141,93	7,16995	0,000182
weight	0,49289	0,274623	2,2	1,21	1,79480	0,115762
origin	0,60216	0,222005	1341,3	494,50	2,71236	0,030096

Table 1.. Multivariate regression model for all explanatory variables

Then these data were investigated for multicollinearity by calculating the eigenvalues:

Value number	Eigenvalues of correlation matrix, and related statistic Active variables only Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	3,616382	45,20478	3,616382	45,2048
2	1,838890	22,98613	5,455273	68,1909
3	1,524847	19,06059	6,980120	87,2515
4	0,566622	7,08277	7,546742	94,3343
5	0,240284	3,00355	7,787026	97,3378
6	0,129920	1,62400	7,916946	98,9618
7	0,055080	0,68850	7,972026	99,6503
8	0,027974	0,34967	8,000000	100,0000

Table 2. Contingency coefficients

$\lambda_{max}/\lambda_{min}$ =129,2766=129,3

I've got a large value, so multicollinearity is present.

For regression model creating were deleted all insignificant variables from the regression equation.

Regression Summary for Dependent Variable: price (CarData1Lab)
R= ,89577150 R?= ,80240658 Adjusted R?= ,77200760
F(2,13)=26,396 p<,00003 Std.Error of estimate: 614,05

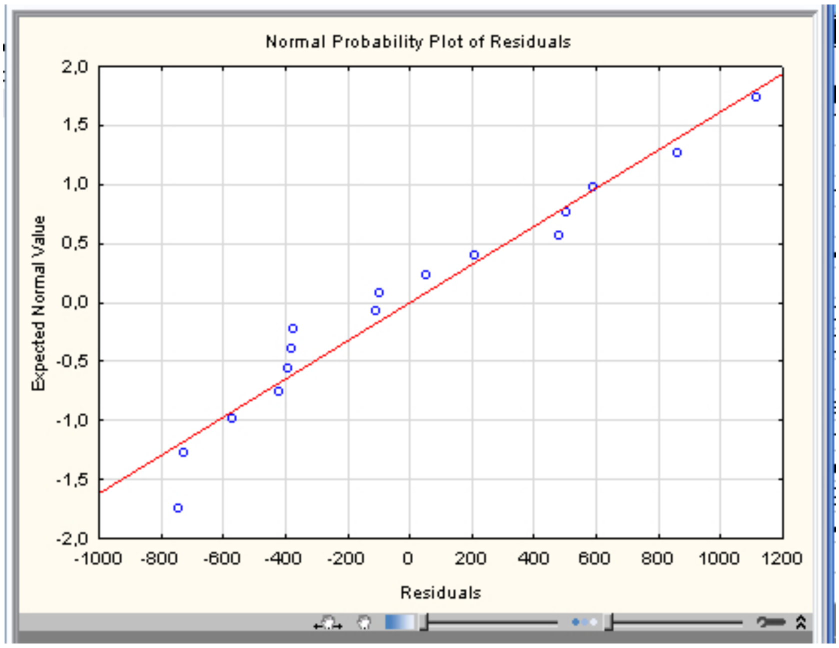
N=16	b*	Std.Err. of b*	b	Std.Err. of b	t(13)	p-value
Intercept			-69849,8	10152,00	-6,88040	0,000011
year	0,938266	0,129606	920,9	127,21	7,23937	0,000007
origin	0,365736	0,129606	814,6	288,69	2,82191	0,014412

Table 3. Multivariate regression model for significant explanatory variables

Were gotten multivariate regression equation: price=920,9* year+ 814,6*origin -69849,8

If $F_p > F_{T[1-q;l;N-(k+1)]}$ the hypothesis that the regression coefficients are equal to zero is rejected.

In our case, F_T =19.40. F(2,13)=29,396. the hypothesis that the regression equation is a constant is rejected.



Picture 5. Graphic of residues

Graphic of residues has been created and according to the Darbin-Watson criterion, we make sure that random disturbances are not correlated:

$d > d_U = 1.37$

	Durbin-Watson d (CarData1Lab) and serial correlation of residuals	
	Durbin-Watson d	Serial Corr.
Estimate	2,021283	-0,026479

Table 4. The Darbin-Watson criterion