# Quantitative Structure-Activity Relationships (QSARs) Tutorial

Single and Multi Linear Regression





This tutorial will show you how to perform a QSAR with MLR and PLS statistical tools

In the Gnumerics file enclosed (Tutorial Data) you will find 29 sugars with different sucrose relative power and 8 different descriptors as reported in the publication. The objective of this tutorial is to make you use of simple informatics tools to establish QSAR models.

**RESEARCH ARTICLE** 

R.K. Singh, M.A. Khan and P.P. Singh, S. Afr. J. Chem., 2014, 67, 12–20, <a href="http://journals.sabinet.co.za/sajchem/">http://journals.sabinet.co.za/sajchem/</a>>

#### Rating of Sweetness by Molar Refractivity and Ionization Potential: QSAR Study of Sucrose and Guanidine Derivatives

Rajesh K. Singh\*, Mohd. A. Khan and Pashupati P. Singh

Department of Chemistry, M.L.K. P.G. College, Balrampur, U.P., India.

Received 31 December 2012, revised 29 August 2013, accepted 19 December 2013.

#### ABSTRACT

A quantitative structure activity relationship study of 31 sucrose derivatives and 30 guanifine derivatives has been undertaken. Their sweetness values, relative to sucrose (RS), have been taken from literature. The study has been made with the help of CAChe Pro software by using eight descriptors, *viz*. electron affinity, ionization potential, electrophilicity index, total energy, heat of formation, steric energy, molar refractivity and solvent accessible surface area. Multi-linear regression (MLR) analysis has been performed with different combinations of descriptors and the quality of regression has been adjudged by the correlation coefficient, cross-validation coefficient and other statistical parameters like the standard error, standard error of the estimate, degrees of freedom, etc. The study indicates that ionization potential appears an important descriptor for sucrose derivatives, whereas molar refractivity appears an important descriptor for guanidine derivatives. The ionization potential alone and in combination with the electrophilicity index, molar refractivity and solvent accessibility surface area provide dependable QSAR models for sucrose derivatives. Molar refractivity alone and in combination with solvent accessibility surface area provide dependable QSAR models for sucrose derivatives. Molar refractivity alone and in combination with solvent accessibility surface area provide solvest neess values obtained by these QSAR models are close to observed sweetness.

QSAR

Pagina 2

12

The tutorial rely on the data available from Singh's paper. You can fine the publication in elearning website

🚺 Download			- 🗆 🗵
Goo ↓ • rigno • Download •	👻 💽 🔽 Cerca Downlo	ad	2
File Modifica Visualizza Strumenti ?			
Organizza 🔻 📆 Apri 🛛 Condividi con 👻	Nuova cartella		
🔆 Preferiti	Nome ^	Ultima modifica	Тіро
🧾 Desktop	QSARToolbox3.3.5Jul2015	7/22/2015 10:28 AM	Cartella d
🐌 Download	a-pdf-mg.exe	9/15/2015 2:47 PM	Applicazic
🔛 Risorse recenti	🙀 FileFormatConverters.exe	6/15/2015 11:18 AM	Applicazic
Lavoro_Univ	SnumericPortable_1,12.17.paf.exe	12/2/2015 7:28 AM	Applicazio
Universita	獨 MWSnap300.exe V	12/9/2015 8:15 AM	Applicazic
Laureandi	G OpenBabel2.3.2a_Windows_Installer.exe	11/23/2015 2:50 PM	Applicazic
	pdfsam-v2_2_4.msi	9/15/2015 1:24 PM	Pacchettc
MieiDocumenti	🚱 picpick_inst.exe	12/9/2015 8:25 AM	Applicazic
Google Drive	🔁 Wirth. 1.pdf	2/23/2015 10:10 AM	Documen <sup>*</sup>
	🔁 Wirth. 2. 1.pdf	2/23/2015 10:47 AM	Document
潯 Raccolte	🔁 Wirth.2.2.pdf	2/23/2015 11:07 AM	Document
Documenti	🔁 Wirth. 2. pdf	2/23/2015 10:26 AM	Document
📄 Immagini			
J Musica			
Video	<b>v</b> 4		Þ
GnumericPortable_1.12.17.paf Applicazione	exe Ultima modifica: 12/2/2015 7:28 AM Data creazione: 12/2/2015 7:29 AM Dimensione: 20.9 MB		

QSAR

Pagina 3

The gnumeric program will be used.

The program can be searched in the net as it is freely available.

Anyway it is also available from elarning.

Download it and install by double click on the file.



#### Answer to the question during the installation as shown in the slides



QSAR

Pagina 5



📙 GnumericPortable	_	미지
Cerca GnumericPortable - 🗸 🐨 🗸 Cerca GnumericPortable		- 2
File       Modifica       Visualizza       Strumenti       ?         Organizza	•	•
Nome     Image: Computer     Image: Computer <td></td> <td>•</td>		•
GnumericPortable.exe Ultima modifica: 12/6/2013 8:24 PM Data creazione: 12/6/2013 8:24 PM Data creazione: 12/6/2013 8:24 PM		

QSAR

Pagina 7

Look for the folder where gnumeric has been installed and open up it

### **Open the file**

w w From Template	Ctrl+N	¢ 🔻	2	f(x)	2 20	10	0% ▼					
oen N	Ctrl+O ⊐ 8	== Q	1 26	•		+==	Až Az	-	d₁ <b>▼</b>	А 👻		
ive W	Ctrl+S	a 🗐										
ive As	Shift+Ctrl+S											
ge Setup												
nt Area & Breaks	•	F	G	н	1	1	К	L	м	N	0	
nt Preview			-									
nt	Ctrl+P											
nd To												
cument Properties												
ll History												
	Ctrl+w											
iit	Ctrl+O											
<u>.</u>											F	
t1 Sheet2 Sheet3			Ope	n a file				SI	um = 0			
												_

By mean of the «file» menu ...

#### **Open the file**



From elearning download the file Rajesh.Student.gnumeric and open it with gnumeric

#### **Initial Table**

			▼ 10	• A A			<b>**</b>		š • 🐝		) 🖅   E	8 -	-
К13		🍨 🐰 🍯	· = [										
	A	В	С	D	E	F	G	н	I.	J	к	L	
1	Dep	oendent Va			Phy	sicochem	ical Desc	riptors					*
2	#	RS	EA	IP	W	MR	SASA	TE	DH	SE			
3	2	1.9999	0.557	5.592	1.878	71.825	135.699	-1682.344	-440.496	117.801			
4	3	3.9994	0.502	5.619	1.830	77.927	152.919	-2451.019	-363.996	114.573			
5	4	4.9992	0.421	5.596	1.749	71.825	137.370	-1682.335	-441.138	121.853			
6	5	4.9992	0.688	5.586	2.009	74.876	140.247	-2066.684	-400.947	116.732			
7	6	19.9986	0.508	5,723	1.861	71.825	135.589	-1682.336	-438.212	122.617			
8	7	19.9986	-0.175	5.614	1.277	71.825	133.584	-1682.342	-438.670	126.687			
9	8	24.9977	0.329	5.871	1.734	77.927	152.595	-2451.015	-359.039	114.619			
10	9	29.9985	0.736	5.888	2.129	74.876	142.065	-2066.680	-401.520	114.594			
11	10	39.9945	-0.321	5.976	1.269	63.698	131.214	-1370.073	-479.375	111.941			
12	11	49.9919	0.337	5.728	1.706	74.876	139.270	-2066.688	-402.632	117.804			
13	12	75.9976	0.417	5.886	1.816	74.876	139.818	-2066.690	-399.686	118.920		1	
14	13	100.0000	0.865	5.891	2.270	77.927	146.285	-2451.029	-362.811	108.918		1	
15	14	119.9776	0.559	5.790	1.926	74.876	140.650	-2066.681	-402.218	113.527			
16	15	119.9776	1.430	5.852	2.997	102.919	183.213	-21831.257	-269.022	101.888			
17	16	149.9685	0.453	6.018	1.881	76.830	145.349	-2375.806	-323.533	91.919			
18	17	159.9926	0.792	5.833	2.177	77.927	145.483	-2451.031	-365.015	107.813			
19	18	199.9862	0.597	5.989	2.011	80.978	157.941	-2835.362	-324.988	110.836			
20	19	219.9885	0.737	5.911	2.135	77.927	147.355	-2451.027	-365.530	105.431			
21	20	299.9853	0.387	5.923	1.798	82.678	152.621	-2490.338	-358.418	107.722			
22	21	374.9730	1.286	6.009	2.817	80.911	148.479	-4564.716	-353.596	108.281			
23	22	399.9447	0.391	5.924	1.802	76.383	146.250	-2375.798	-323.364	107.853			_
24	24	649.9800	0.494	6.020	1.919	77.927	145.051	-2451.034	-363.678	109.790			
25	25	799.8343	1.110	5.973	2.580	83.894	149.738	-6678.402	-341.980	108.112			
26	26	799.8343	1.250	5.920	2.751	86.878	156.431	-8792.080	-331.894	106.298			
27	27	1000.0000	0.543	6.115	1.989	76.235	143.952	-2475.061	-364.739	99.798			
28	28	2199.8851	0.912	6.105	2.370	80.978	151.349	-2835.376	-326.350	99.799			
29	29	2999.8532	1.694	6.031	3.441	83.962	155.358	-4949.061	-315.982	98.052			
30	30	/498.9421	1.713	6.054	3.473	89.309	162.623	-9295.451	-296.468	97.150			
31	- 31	/498.9421	1.618	5.974	5.308	92.912	166.505	-11290.109	-284.652	95.312			
22	-												+
34													
35													
35													
50	1												
	Lisboot2	Charaka											<u> </u>

The opened file should appear as in the slide.

There is a dependent variable (RS = relative sweetness) and 8 descriptors

#### **Descriptors**

Molar refractivity is calculated by the Lorenz-Lorentz formula

 $MR = \frac{n^2 - 1}{n^2 + 1} \times \frac{M}{\rho},$ 

Parr *et al.* introduced the electrophilicity index ( $\omega$ ) in terms of the chemical potential and hardness.<sup>30</sup> The operational definition of the electrophilicity index may be written as,

 $\omega = \mu^2/2\eta$ . W in the table

The total energy (TE) of a molecular system is the sum of the total electronic energy (Eee) and the energy of internuclear repulsion (Enr).<sup>31</sup>

TE = Eee + Enr.

The solvent accessibility surface area (SASA) is the surface area of a biomolecule that is accessible to a solvent and is usually quoted in square angstroms. Lee and Richards first described the solvent accessible surface area (SASA) of a molecular surface.<sup>34</sup>

QSAR

The heat of formation is defined as:

$$\label{eq:deltaHf} \begin{split} \Delta H_f &= E_{elect\,+}\,E_{nuc} - E_{isol}\,+\,E_{atom}\,,\\ & \text{DH in the table} \end{split}$$

According to the Koopman's theorem, the ionization potential is simply the eigenvalue of the highest occupied molecular orbital (HOMO) with change of sign and the electron affinity is the eigenvalue of the lowest unoccupied molecular orbital (LUMO) with change of sign.<sup>29</sup>

#### EA and IP in the table

The steric energy of a molecule is the sum of the molecular mechanics potential energies calculated for the bonds, bond angles, dihedral angles, nonbonded atoms and so forth. It is specific to mechanics and depends upon the force-field used.<sup>33</sup>

#### SE in the table

Pagina 11

Here are described the 8 used descriptors

3	Undo 🤅		Ctrl+Z	🧄 👻		9 E .	f(x) 8. z. Miz Miz M	a   <b>36</b> 1009	*				
Fimes	Redo		Ctrl+Y	- A 7			<b>H D</b>	88 9 9	s . 🕬			-	
12	Cut		Ctrl+X	s									-
	Dacte		Ctrl+V		E	E	G	ш			r		-
	Paste sper		hift+Ctrl+V	U	E Dhy	r sicochem	ical Desi	rintors		J	N		٩
2				TD	W	MR	SASA	TE	DH	SE			
	Clear		•	E 502	1 070	71.005	125 600	1692.244	440.406	117.001			
4 8	<u>D</u> elete		+	5.610	1.878	77.027	152.099	-2451.010	-262.006	11/.801			
5 2	Modify		•	5 506	1.830	71.927	132.919	-1692 325	-441 129	121 952			
6				5 586	2.009	74.876	140 247	-2066 684	-400 947	116 732			
7	Sgarch		Ctrl+F	5 723	1.861	71.825	135 589	-1682 336	-438 212	122 617			
8	Search & R	teplace	Ctrl+H	5.614	1.277	71.825	133.584	-1682.342	-438.670	126.687			
9	Choot		•	5 871	1 734	77 927	152 595	-2451 015	-359.039	114 619			
10	Select		(	5 888	2 1 2 9	74 876	142 065	-2066680	-401520	114 594			
11	Recalculati			5.976	1.269	63.698	131.214	-1370.073	-479.375	111.941			
12 -				5,728	1.706	74.876	139,270	-2066,688	-402.632	117,804			
18 2	Preference	2S		5.886	1.816	74,876	139.818	-2066.690	-399.686	118,920			
14	13	100.0000	0.865	5.891	2.270	77.927	146.285	-2451.029	-362.811	108.918			
15	14	119.9776	0.559	5.790	1.926	74.876	140.650	-2066.681	-402.218	113.527			
16	15	119.9776	1.430	5.852	2.997	102.919	183.213	-21831.257	-269.022	101.888			
17	16	149.9685	0.453	6.018	1.881	76.830	145.349	-2375.806	-323.533	91.919			
18	17	159.9926	0.792	5.833	2.177	77.927	145.483	-2451.031	-365.015	107.813			
19	18	199.9862	0.597	5.989	2.011	80.978	157.941	-2835.362	-324.988	110.836			
20	19	219.9885	0.737	5.911	2.135	77.927	147.355	-2451.027	-365.530	105.431			
21	20	299.9853	0.387	5.923	1.798	82.678	152.621	-2490.338	-358.418	107.722			
22	21	374.9730	1.286	6.009	2.817	80.911	148.479	-4564.716	-353.596	108.281			
23	22	399.9447	0.391	5.924	1.802	76.383	146.250	-2375.798	-323.364	107.853			
24	24	649.9800	0.494	6.020	1.919	77.927	145.051	-2451.034	-363.678	109.790			
25	25	799.8343	1.110	5.973	2.580	83.894	149.738	-6678.402	-341.980	108.112			
20	26	799.8343	1.250	5.920	2.751	86.878	156.431	-8/92.080	-351.894	106.298			
20	27	2100.0000	0.545	0.115	1.989	/0.235	145.952	-24/5.061	-304.739	99.798			
20	28	2199.8851	1.604	6.021	2.570	80.978	151.549	-2835.376	-320.350	99.799			
30	29	7408 0421	1.094	6.051	2.472	83.902	162,622	-4949.061	-315.982	98.052			
31	21	7498.9421	1./15	5.074	2 209	02.012	166 505	-11200 100	-290.408	97.150			
32	- 21	7420.9421	1.018	5.974	5.508	52.912	100.303	11290.109	204.032	95.512			
33													
34													
35													
36	1												
	4												(p
nitial Tab	ale Sheet2	Sheet3			Comercial								-

QSAR

Pagina 12

The first thing to do is to check the linear distribution of the dependent variable RS. So copy the column of data belonging to RS it in the memory.

ns			- 10 -	AA	A		P = 8	: Q9 %	5 . <b>%</b>	I		⊞ ▼	•
	4	> 💥 🖑 👻	= [										
Α		в	С	D	E	F	G	н	1	J	к	L	
1													-
2		cu <u>r</u>											
3	1.5	<u>C</u> opy											
4		<u>P</u> aste											
5		Paste Special											
7		Insert 1 Cel											
0	8	Delete 1 Cell										-	
9		Clear Contents											
10	-	clear conterns											
11	Ø	Add Comment											
12		Add Hyperlink											
13	-		_										
14	2	Eormat 1 Cell											
15		Cell	•										
16		Column	•										
17		Row	•										
18													
19													
20													
21													
22													
23													
25				1									
26													
27				1									
28													
29													
30													
31													
32													
33													
34													
26													
30													
1									_				-

QSAR

Pagina 13

.... and paste it the into a new sheet

🐻 * Raj	esh.Teacher.gr	umeric - Gnum	eric		-							_10	×
Ele E	dit <u>V</u> iew Inse	rt F <u>o</u> rmat <u>T</u> oo	ls <u>S</u> tatistics	: <u>D</u> ata <u>H</u> el;	þ								
n r		& I 🔏 🗈	n   🤞	- 📣	- 9	$\sum f(\mathbf{x})$	n. z.	100%	-				
								1.000 .00					
Times Ne	w Roman	· · · ·	12 •				y Lu aa	9 %	• *****		<b>E</b>   E	·	×
A1	4	X 🖑 🔹	= RS										
	Α	В	С	D	E	F	G	н	1	J	К	L	
1	RS												~
2	1.9999												_
3	3.9994												
4	4.9992												_
5	4.9992												_
6	19.9986												- 1
/	19.9986												-
°	24.9977												-
10	29.9985												- 1
11	49 9919												-
12	75 9976												-
13	100.0000												- 1
14	119 9776												-
15	119.9776												
16	149,9685												
17	159.9926												
18	199.9862												
19	219.9885												
20	299.9853												
21	374.9730												
22	399.9447												
23	649.9800												
24	799.8343												
25	799.8343												
26	1000.0000												
27	2199.8851												
28	2999.8532												
29	7498.9421												
30	7498.9421												
31	3												
Initial Tal	ole Sheet2 She	et3							Sur	m=25869.04	01937203		
									1			- T	111

QSAR

Pagina 14

) 🖻 🛛	<b>.</b>   <b>.</b>	C <u>e</u> ls	Ctrl++	ø -	-   🍛	$\sum f(\mathbf{x})$	z. Z Ma	100%	•				
imes New R	oman + 🗄	Rows		A A			. 33	69 %	. 🎭		<b>₽</b>   ⊞	-	-
		Sheet									01		_
<b>\1</b>		Chart											
	A	Image			E	F	G	н	1.1	J	К	L	
1	RS	Eurotion											<u>^</u>
2	1.999	Eurotion Wrapper											
3	3.999	Name	E2										_
4	4.99	Commont	15										_
5	4.99! 9	Comment											_
0	19.99	nyperiink	Ctrl+K										_
·	24.0077	Special											_
0	24.5911			-									
10	39 9945												
11	49 9919												
12	75 9976												
13	00.0000												
14	19.9776												
15 1	19.9776												
16 1	149.9685												
17 1	159.9926												
18 1	199.9862												
19 2	219.9885												
20 2	299.9853												
21	374.9730												
22	399.9447												_
23 6	549.9800												_
24	799.8343												_
25	000 000												_
20 1	100.0000												_
27 2	199.865.	2											_
20 2	498 942	1											
30 7	498.942	1											
31	100.074			-									-
1													Þ
nitial Table	Sheet2 S	heet3		l	nsert a Cha	rt			Su	m=25869.040	01937203		11

Pagina 15

To check for linearity, just insert a column type chart

QSAR



QSAR

Pagina 16

As you can see, the preview is showing the plot

e E	dit ⊻jew Insert	Format To	ols Statistics	Data Hel	0				,				
	8 🗖   占 🕴	l 🖌 🗆	) 🖸   🥎	- 🔶	-   🎱	$\sum f(x)$	8, 2, ≌2 288	100%	~				
imes Ne	w Roman		12 -			🛛 🗍 E	8 🗆 22	9 %	× 👘 0		🖲   E	*	•
1	۵	* 🖉 👻	= RS										
	A	В	С	D	E	F	G	Н	1	J	К	L	
1	RS												4
2	1.9999												
3	3.9994												
4	4.9992												
5	4.9992												
6	19.9986		9					ρ					
7	19.9986		800	0.0000	_								
8	24.9977		700	0.0000-									
9	29.9985		600	0.0000 -									
10	39.9945		500	0.0000 -									
11	49.9919		400	0000									
12	75.9976		9 ***					P					
13	100.0000		300	0.0000 -									
14	119.9776		200	- 0000.0									
15	119.9776		100	0.0000 -			- 0						
16	149.9685			0.0000			տորդովվ						
17	159.9926			2	6	10 14	18 22	26					
18	199.9862		Ŭ			Ŭ		Ŭ					
19	219.9885												
20	299.9853												
21	374.9730												
22	399.9447												
23	649.9800												
24	799.8343												
25	799.8343												
26	1000.0000												
27	2199.8851												
28	2999.8532												
29	7498.9421												
30	7498.9421												
31													_
22	र												

QSAR

Pagina 17

Then click in the area of the sheet grid and the plot will appear. Clearly there is no linear distribution of the dependent variable

	shareacherigh	umeric - Gn	umeric										_	
le <u>E</u> di	it <u>V</u> iew Inser	t Format	Tools S	tatistics	Data He	þ								
) 🖻		💁   🕌	00	🥎	- 🔶	-   🎱	The first	8. 2. Ma Mie	100%	~				
ns			- 10	•	AA		E	8 🗆 88	9 %	. ⊕.o	00 10	E E	) <b>v</b>	-
	4	💥 🦪	• ==	=log(A	2									
	Α	в		с	D	E	F	G	н	1		к	L	1
1	RS													
2	1.9999									=log(A2	1			
3	3.9994													
4	4.9992													
5	4.9992													
6	19.9986			_										
7	19.9986			800	0.0000	_	_							
8	24.9977			700	0.0000 -									
9	29.9985			600	0.0000 -									
10	39.9945			500	0.0000									
11	49.9919			1										
12	75.9976			400	0.0000-									
13	100.0000			300	0.0000 -									
14	119.9776			200	0.0000 -									
15	119.9776			100	0.0000 -			_						
16	149.9685				0 0000			տոոոՈ						
17	159.9926				2	6	10 14	18 22	26					
18	199.9862													
19	219.9885													
20	299.9853													
21	374.9730													
22	399.9447													
23	649.9800													
24	799.8343													
25	799.8343													
26	1000.0000													
27	2199.8851													
28	2999.8532													
29	7498.9421													
30	7498.9421													
31														
22	•													D E
tial Tabl	le Sheet2 Shee	et3									Sum	=0		Ξ.

QSAR

Pagina 18

Let's transform it in the logarithm scale as shown in the slide

🐻 * Raje	sh.Teacher.g	numeric - Gnum	eric										
Eie Edi	t <u>V</u> iew Inse	rt Format Too	ls <u>Statistics</u>	Data Hel	D								
D 🗅	8 8	🕹   💥 🖸	0   🤞	• 🚸	-   🍛	∑ f(x)	n, z, Miz Min	100%	-				
Sans		•	10 -	AAA				9 %	• 👫		•=   E	•	-
13	4	- 🖏 🗶 🗸	-										
	•	D.		D			6				×		
1	PC	в	C	0	E	F	0			J	~	L	
2	1 9999								0 201				
3	3 9994							Î	0.301				-
4	4 9992							l					-
5	4.9992												$\pm$
6	19.9986		_										+
7	19.9986		800	0.0000									-
8	24.9977		700	- 0000.00									
9	29.9985		600	0.0000 -									
10	39.9945		500	0 0000 -									
11	49.9919												
12	75.9976		400	0.0000 -									
13	100.0000		300	0.0000 -									
14	119.9776		200	0.0000 -									
15	119.9776		100	- 0000.00			- 81						
16	149.9685			0.0000		مممممم	ոորդորվ						
17	159.9926			2	6	10 14	18 22	26					
18	199.9862												
19	219.9885												
20	299.9853												
21	374.9730												
22	399.9447												
23	649.9800												
24	799.8343												
25	799.8343												
26	1000.0000												
27	2199.8851												
28	2999.8532												+
29	7498.9421					-							+
30	7498.9421												+
31	-												<u> </u>
Initial Tabl	e Sheet2 Sh	eet3								Sum	-0		-

QSAR

Pagina 19

				<u>√</u> · ₩			112 11in	1.00%				
ins		•	10 •	AA				9 %	· 900			• •
	4	× 🖉 -	= =lo	g(A2)								
	Α	В	С	D	E	F	G	н	- 1	J	К	L
1	RS									lea(A 20)		<b></b>
2	1.9999								0.301	0g(H29)		
3	3.9994											
4	4.9992											
5	4.9992											
6	19.9986											
7	19.9986			5000.0000 T								
8	24.9977			7000.0000 -								
9	29.9985			5000.0000 -								
10	39.9945			5000.0000 -								
11	49.9919			000 0000								
12	75.9976			+000.0000 -								
13	100.0000			3000.0000 -								
14	119.9776			2000.0000 -								
15	119.9776		:	1000.0000 -								
16	149.9685			0 0000			الاسمم					
17	159.9926				6	10 14	18 22	26				
18	199.9862											
19	219.9885											
20	299.9853											
21	374.9730											
22	399.9447											
23	649.9800											
24	799.8343											
25	799.8343											
26	1000.0000											
27	2199.8851											
28	2999.8532											
29	7498.9421											
30	7498.9421								9			
31												
22												

QSAR

Pagina 20

Then apply the formula for the cell «I2» to the other cells just grabbing the small square in the cell right bottom corner

🐻 * Raje	sh.Teacher.gr	umeric - Gn	umeric										_	
Ele Ed	lit <u>V</u> iew Inser	t Format	<u>T</u> ools <u>S</u>	tatistics D	ata <u>H</u> el	р								
		🗛   🙀	D 🖪	🧄 ·	• 🚸	- 🖌 🎧	$\sum f(\mathbf{x})$		100%	<b>.</b>				
_				<u> </u>					1.000.000					
Sans			▼  10	• A	A				9 %	· *********	ő   +3		•	•
12	4	X 🖑		=log(A2)										
	A	в		с	D	E	F	G	н		1	к	L	
1	RS													-
2	1.9999									0.301				
3	3.9994									0.602				
4	4.9992									0.6989				
5	4.9992									0.6989				
6	19.9986									1.301				
7	19.9986			8000.0	000 7					1.301				
8	24.9977			7000.0	000 -					1.3979				
9	29.9985			6000.0	000 -					1.4771				
10	39.9945			5000.0						1.602				
11	49.9919									1.6989				
12	75.9976			4000.0	1					1.8808				
13	100.0000			3000.0	000 -					2				
14	119.9776			2000.0	000 -					2.0791				
15	119.9776			1000.0	000 -			_		2.0791				
16	149.9685			0.0	000					2.176				
17	159.9926				2	6'	10 14	18 22	26	2.2041				
18	199.9862									2.301				
19	219.9885									2.3424				
20	299.9853									2.4771				
21	374.9730									2.574				
22	399.9447									2.602				
23	649.9800									2.8129				
24	799.8343									2.903				
25	799.8343									2.903				
26	1000.0000									3				
27	2199.8851									3.3424				T
28	2999.8532									3.4771				
29	7498.9421									3.875				
30	7498.9421									3.875				
31														-
22	4													Ŀ
Initial Tab	le Sheet2 She	et3									Sum=61	.9827		1.

QSAR

Pagina 21

And all the data will be transformed



Make the plot as shown before and now the linearity is present.

Sans			▼ 10 ▼	AA	A		i = #	3 🗆 🔐	9%	• • <u>•</u> • • • • • • • • • • • • • • • •	-	•	•	-	
C37		🤹 🐰 🎺 🦷	- =												
	Α	в	С	D	E	F	G	н	1	J	К	L	м		
2	#	RS	EA	IP	W	MR	SASA	TE	DH	SE				-	
3	2	1.9999	0.557	5.592	1.878	71.825	135.699	-1682.344	-440.496	117.801					
4	3	3.9994	0.502	5.619	1.830	77.927	152.919	-2451.019	-363.996	114.573				_	
5	4	4.9992	0.421	5.596	1.749	71.825	137.370	-1682.335	-441.138	121.853				- 11	
6	5	4.9992	0.688	5.586	2.009	74.876	140.247	-2066.684	-400.947	116.732				- 11	
7	6	19.9986	0.508	5.723	1.861	71.825	135.589	-1682.336	-438.212	122.617				- 11	
8	7	19.9986	-0.175	5.614	1.277	71.825	133.584	-1682.342	-438.670	126.687				- 11	
9	8	24.9977	0.329	5.871	1.734	77.927	152.595	-2451.015	-359.039	114.619				- 11	
10	9	29.9985	0.736	5.888	2.129	74.876	142.065	-2066.680	-401.520	114.594				- 11	
11	10	39.9945	-0.321	5.976	1.269	63.698	131.214	-1370.073	-479.375	111.941				- 11	
12	11	49.9919	0.337	5.728	1.706	74.876	139.270	-2066.688	-402.632	117.804				- 11	
13	12	75.9976	0.417	5.886	1.816	74.876	139.818	-2066.690	-399.686	118.920				- 11	
14	13	100.0000	0.865	5.891	2.270	77.927	146.285	-2451.029	-362.811	108.918				- 11	
16	14	119.9776	0.559	5.790	1.926	/4.8/0	140.050	-2066.681	-402.218	101.000				- 10	
17	16	119.9776	1.450	5.852	2.997	76.920	145.215	-21851.257	-209.022	01.010				- 11	
18	17	149.9085	0.455	5 022	2 177	77.027	145.549	-2451.021	-265 015	107 012				- 11	
19	19	100.0862	0.792	5 090	2.177	20.079	145.465	-2431.031	-224 099	110 926				- 10	
20	10	210 0885	0.737	5 011	2.011	77 927	147 355	-2451.027	-365 530	105.431				- 11	
21	20	200 9853	0.387	5 923	1 798	82 678	152 621	-2490 338	-358 418	107 722					
22	21	374 9730	1 286	6 009	2.817	80.911	148 479	-4564 716	-353 596	108 281					
23	22	399 9447	0 391	5 924	1 802	76 383	146 250	-2375 798	-323 364	107 853					
24	24	649 9800	0 494	6 0 2 0	1 9 1 9	77 927	145 051	-2451 034	-363 678	109 790					
25	25	799.8343	1.110	5.973	2.580	83,894	149,738	-6678.402	-341.980	108.112					
26	26	799.8343	1.250	5.920	2.751	86.878	156.431	-8792.080	-331.894	106.298					
27	27	1000.0000	0.543	6.115	1.989	76.235	143.952	-2475.061	-364.739	99.798					
28	28	2199.8851	0.912	6.105	2.370	80.978	151.349	-2835.376	-326.350	99.799					
29	29	2999.8532	1.694	6.031	3.441	83.962	155.358	-4949.061	-315.982	98.052				- 11	
30	30	7498.9421	1.713	6.054	3.473	89.309	162.623	-9295.451	-296.468	97.150				- 11	
31	31	7498.9421	1.618	5.974	3.308	92.912	166.505	-11290.109	-284.652	95.312				- 11	
32															
33															
34														- 11	
35														_	
27			í	1											
20				4										<b>V</b>	
	4													▶	
Initial T	Manag	e sheets									Sum=0			//	
	Incert														
	Appen	d													
	Duper	te													
	Farmon	re la													
	Renam	ie i													Pagina 2
	Resize														Fagina 2
	0.1														

Duplicate the sheet by right clicking on the sheet name.

2		Energinate Toola	Statistic	s <u>D</u> ata	Help								
18		🖴 🚨   🎽 🕞		•	(† -		∑ f( <b>x</b> )	🗽 🐝 🛛 🛍	100%				
nes New	Roman	•	12 -	A				3 🗆 88	<b>9</b> % ·	- <b>*</b> 2 _	8 🔲 🗉	E E	-
			- 1	D	-	-	6				K		
1		nendent Variable	C		 D	F hysicocl	hemical	Descriptor		J	N	L	Ivi
2	#	RS	FΔ	ID	w	MR	SASA	TE	DH	SE			
3	2	1 0000	0.557	5 502	1 0 70	71 925	125 600	-1692 244	-140.406	117 201			
4	2	2 0004	0.557	5.592	1.878	77.027	152.099	-2451.010	-262.006	114 572			
5	3	4 0002	0.302	5 506	1.850	71.927	127 270	-1692 225	-441 120	121 052			
6	5	4,9992	0.421	5 596	2,000	74.076	140.247	-2066 694	-400.047	116 722			
7	6	10.00%	0.000	5 722	1 961	71 025	125 590	-1602 226	-420 212	122 617			
0	7	10,0006	0.508	5.725	1.801	71.825	122 504	1602.330	420 670	122.017			
Q I	0	24.0077	0.220	5 071	1.277	77.027	152 505	-2451.015	-250.020	114 610			
10	0	24.9977	0.529	5.8/1	2 120	74.921	142.065	-2451.015	-401 520	114.019			
11	10	29.9963	-0.221	5.076	1 260	62 609	121 214	-1270.072	-470.275	111.041			
12	11	40.0010	0.321	5 720	1.209	74 976	120 270	-2066 699	-402.622	117 004			
12	12	49.9919	0.337	5.728	1.016	74.870	120 010	-2000.088	200.696	117.804			
14	12	100,0000	0.417	5.880	1.810	77.027	146 205	-2066.690	-399.080	108.920			
15	13	110.0000	0.805	5.891	1.026	74.927	140.285	2451.029	402.210	112 527			
16	14	119.9776	1.420	5.790	1.920	102.010	140.050	-2000.081	-402.218	101.000			
17	15	119.9776	0.452	5.852	2.997	76.020	145 240	2275 006	202.522	01.010			
10	10	149.9085	0.455	5.022	1.881	70.830	145.549	2451 021	265 015	107.012			
10	10	100.0960	0.792	5.833	2.1//	20.070	145,485	2431.031	-305.015	110.026			
20	10	210.0005	0.397	5.989	2.011	30.978	137.941	-2855.502	265 520	105 421			
21	19	219.9885	0.757	5.911	2.135	02.670	147.555	2400 220	250 410	107.722			
22	20	299.9855	1.206	5.923	1.798	82.078	132.021	4564 716	252 506	107.722			
22	21	200.0447	0.201	5.009	2.817	76 202	146.479	-4304.710	202.264	108.281			
24	22	599.9447	0.391	5.924	1.802	70.383	146.250	-25/5./98	262 670	107.855			
25	24	700 9242	1 1 1 1 0	5 072	2.590	02 004	140 720	-6679 402	-241 000	109.790			
26	26	700 9242	1.110	5.973	2.380	06 070	156 421	-0078.402	-221 004	106,112			
27	20	1000 0000	0.542	6 115	2.751	76 225	142.052	-2475.061	-264 720	00 709			
28	20	2100.0000	0.012	6 105	2 270	20.070	151 240	-2925 276	-226 250	99.798			
29	20	2177.0821	1 604	6.021	2.370	83.062	155 350	-4040.041	-315 002	08 052			
30	30	7408 0421	1 712	6.054	3 472	80 300	162 622	-0205 451	-206 140	97.150			
31	31	7/08 0/21	1.619	5 074	3 300	02 012	166 505	-11200 100	-284 652	05 312			
32	21	/+20.2421	1.018	5.914	5.308	94.912	100.505	-11290.109	-204.002	95.512			
33													
34													
35													
36													
27													

QSAR

Pagina 24

Rename the sheet into «New Table»



Highlight the transformed RS variable and copy into the memory with the right click

Times New	Roman	-	12 -	AA	A		3 2		9 % ·	• • …		9 🖽	•	-
B3		<li>- 1%</li>	= 1.999	19										
	A		С	D	E	F	G	н	1		к	L	M	
1	Dependen	t Variabl	e		Р	hysicoc	hemical	Descriptor	5					<u>_</u>
2	# R	S	EA	IP	W	MR	SASA	TE	DH	SE				
3	2 1.99	199	0.557	5.592	1.878	71.825	135.699	-1682.344	-440.496	117.801				
4	3 3.99	94	0 502	5 619	1.830	77.927	152.919	-2451.019	-363.996	114.573				_
5	4 4.9	🕌 Cu <u>t</u>		596	1.749	71.825	137.370	-1682.335	-441.138	121.853				-
0	5 4.9	🕒 Сору		286	2.009	74.876	140.247	-2066.684	-400.947	116.732				
8	7 19.	Paste		614	1.801	71.825	122 594	-1682.330	-438.212	122.01/				-
9	8 244	Paste S	peda	871	1 734	77 927	152 595	-2451 015	-359 039	114 619				
10	9 29.5	Insert 2	9 Cells	888	2.129	74,876	142.065	-2066.680	-401.520	114.594				
11	10 39.9	Delete 2	29 Cells	976	1.269	63.698	131.214	-1370.073	-479.375	111.941				
12	11 49.9	A Clear Cr	ontents	728	1.706	74.876	139.270	-2066.688	-402.632	117.804				_
13	12 75.9		Jucino -	886	1.816	74.876	139.818	-2066.690	-399.686	118.920				_
14	13 100.	Add Cor	ment	891	2.270	77.927	146.285	-2451.029	-362.811	108.918				-
15	14 119.	Add Hyp	perlink	/90	1.926	/4.8/6	140.650	-2066.681	-402.218	101.000				-
17	16 149			012	1 001	76.830	145 240	-21851.257	-209.022	01 010				
18	17 159	Eormat:	29 Cells	. 833	2 177	77 927	145 483	-2451.031	-365.015	107 813				
19	18 199.	Celum		989	2.011	80,978	157,941	-2835,362	-324,988	110.836				
20	19 219.	Pow		911	2.135	77.927	147.355	-2451.027	-365.530	105.431				
21	20 299.:	000	0.307	5.923	1.798	82.678	152.621	-2490.338	-358.418	107.722				
22	21 374.9	730	1.286	6.009	2.817	80.911	148.479	-4564.716	-353.596	108.281				
23	22 399.9	447	0.391	5.924	1.802	76.383	146.250	-2375.798	-323.364	107.853				
25	24 049.5	2/2	1 1 1 1 0	5.072	2.500	92 904	145.051	-2451.034	-241.090	109.790				
26	26 799.5	1343	1.110	5 920	2.580	86 878	156 431	-8792.080	-331 894	106 298				-
27	27 1000	0000	0.543	6.115	1.989	76.235	143.952	-2475.061	-364.739	99.798				
28	28 2199	8851	0.912	6.105	2.370	80.978	151.349	-2835.376	-326.350	99.799				
29	29 2999.	8532	1.694	6.031	3.441	83.962	155.358	-4949.061	-315.982	98.052				
30	30 7498.	9421	1.713	6.054	3.473	89.309	162.623	-9295.451	-296.468	97.150				_
31	31 7498.	9421	1.618	5.974	3.308	92.912	166.505	-11290.109	-284.652	95.312				-
32			-											-
34			-											-
35			-											-
36														
27	-												1	
		1.								[				

Then paste it to the «New Table» sheet. Use the «paste special» option

100 ° R	ajesh.1	eacher.gnumeric - Gnumer	ic										<u>_                                    </u>
Ele	East 1	njew insert ⊢ormat ioois ∎ I D D D I V D	Statistic	s Data	Heb		S. 11.1	8, 2, 6,	1008/				
							264 J 147	31≵ 348   888	¢ 100 %				
Times	New Ro	nan 🔻	12 🔻	A					9 % ·	900 m			· •
B3		- 🦫 🐇 🥥	= 1.999	99									
	A	В	С	D	E	F	G	н	1	J	к	L	м
1		Dependent Variable			Р	hysicoc	hemical	Descriptor	s				-
2	#	RS	EA	IP	W	MR	SASA	TE	DH	SE			
3	2	1.9999	0.557	5.592	1.878	71.825	135.699	-1682.344	-440.496	117.801			
4	3	3.9994	0.502	5.619	1.830	77.927	152.919	-2451.019	-363.996	114.573			
5	4	4.9992	0.421	5.596	1.749	71.825	137.370	-1682.335	-441.138	121.853			
6	5	4.9992	0.688	5.586	2.009	74.876	140.247	-2066.684	-400.947	116.732			
/	6	19.9986	0.508	5.723	1.861	71.825	135.589	-1682.336	-438.212	122.617			
8	- /	19.9986	-0.175	5.614	1.277	71.825	133.584	-1682.342	-438.670	126.687			
- 10	8	24.9977	0.329	5.8/1	1.734	71.927	152.595	-2451.015	-359.039	114.619			
11	10	29.9985	0.730	5.888	2.129	/4.8/0	142.005	-2000.080	-401.520	114.594			
12	11	40.0010	0.2	Paste 9	special					× 204			
13	12	75 0076	0.5	Paste tv	pe	( Cell o	operation	Region of	peration	1 2 920			
14	13	100,0000	0.8	O AL			None	None		8.918			
15	14	119 9776	0.5	0					-	527			
16	15	119,9776	1.4	0.00	nen.		Mūu	O Iran	ispose	888			
17	16	149,9685	0.4	As	Value		Subtract	C Final	Jorizontally	.919			
18	17	159,9926	0.7	O Eor	mats	0	Multiply		nonzontally	7.813			
19	18	199.9862	0.5	0.00	mente		Divide	C Flip	Vertically	0.836			
20	19	219.9885	0.7 -				of the state			5.431			
21	20	299.9853	0.3 F	Skip Bl	anks 🥅		ange formula	e		7.722			
22	21	374.9730	1.2							8.281			
23	22	399.9447	0.3	Hel	p	Paste L	ink	Cancel	<u>o</u> k	7.853			
24	24	649.9800	0.4							□ <u>₽.790</u>			
25	25	799.8343	1.110	5.975	2.580	83.894	149.738	-00/8.402	-341.980	108.112			
20	26	799.8343	1.250	5.920	2.751	86.878	156.431	-8792.080	-331.894	106.298			
2/	2/	1000.0000	0.543	6.115	1.989	76.235	143.952	-24/5.061	-364.739	99.798			
20	28	2199.8851	1.604	6.021	2.570	80.978	151.349	-2835.370	-320.330	99.799			
30	29	7409 0421	1.094	6.051	2 472	85.902	162,622	-4949.001	-206.469	98.052			
31	31	7498.9421	1.619	5 974	3 308	02 012	166 505	-11200 100	-290.408	97.130			
32	- 31	/4/0.7421	1.018	5.974	5.508	52.912	100.505	11290.109	204.052	55.514			
33				-									
34													
35													
36													=1
27	्												
Initial	Table N	lew Table Sheet2 Sheet3									Sum=1.99	99	

QSAR

Pagina 27

... and select «as value»

Check	dependent	variable	(RS)

	3	]   🖴 📥   🔏 🗔		•	<u>ج</u>	2	$\sum f(\mathbf{x})$	🔓 z 🙀 🗎 🛍	100%				
Sans			10 👻	A	<u>A</u>		] 🗏 🆽		9% ·			•	•
G33		🤹 💥 🐗 👻	=										
	Α	В	С	D	E	F	G	Н	I.	J	К	L	
1		Dependent Variable			Р	hysicoc	hemical	Descriptor	5				
2	#	RS	EA	IP	W	MR	SASA	TE	DH	SE			
3	2	0.3010	0.557	5.592	1.878	71.825	135.699	-1682.344	-440.496	117.801			
4	3	0.6020	0.502	5.619	1.830	77.927	152.919	-2451.019	-363.996	114.573			
5	4	0.6989	0.421	5.596	1.749	71.825	137.370	-1682.335	-441.138	121.853			
6	5	0.6989	0.688	5.586	2.009	74.876	140.247	-2066.684	-400.947	116.732			
7	6	1.3010 ഹ	0.508	5.723	1.861	71.825	135.589	-1682.336	-438.212	122.617			
8	7	1.3010	-0.175	5.614	1.277	71.825	133.584	-1682.342	-438.670	126.687			
9	8	1.3979	0.329	5.871	1.734	77.927	152.595	-2451.015	-359.039	114.619			
10	9	1.4771	0.736	5.888	2.129	74.876	142.065	-2066.680	-401.520	114.594			
11	10	1.6020	-0.321	5.976	1.269	63.698	131.214	-1370.073	-479.375	111.941			
12	11	1.6989	0.337	5.728	1.706	74.876	139.270	-2066.688	-402.632	117.804			
13	12	1.8808	0.417	5.886	1.816	74.876	139.818	-2066.690	-399.686	118.920			
14	13	2.0000	0.865	5.891	2.270	77.927	146.285	-2451.029	-362.811	108.918			
15	14	2.0791	0.559	5.790	1.926	74.876	140.650	-2066.681	-402.218	113.527			
16	15	2.0791	1.430	5.852	2.997	102.919	183.213	-21831.257	-269.022	101.888			
17	16	2.1760	0.453	6.018	1.881	76.830	145.349	-2375.806	-323.533	91.919			
18	17	2.2041	0.792	5.833	2.177	77.927	145.483	-2451.031	-365.015	107.813			
19	18	2.3010	0.597	5.989	2.011	80.978	157.941	-2835.362	-324.988	110.836			
20	19	2.3424	0.737	5.911	2.135	77.927	147.355	-2451.027	-365.530	105.431			
21	20	2.4771	0.387	5.923	1.798	82.678	152.621	-2490.338	-358.418	107.722			
22	21	2.5740	1.286	6.009	2.817	80.911	148.479	-4564.716	-353.596	108.281			
23	22	2.6020	0.391	5.924	1.802	76.383	146.250	-2375.798	-323.364	107.853			
24	24	2.8129	0.494	6.020	1.919	77.927	145.051	-2451.034	-363.678	109.790			_
25	25	2.9030	1.110	5.973	2.580	83.894	149.738	-6678.402	-341.980	108.112			_
26	26	2.9030	1.250	5.920	2.751	86.878	156.431	-8792.080	-331.894	106.298			_
27	27	3.0000	0.543	6.115	1.989	76.235	143.952	-2475.061	-364.739	99.798			_
28	28	3.3424	0.912	6.105	2.370	80.978	151.349	-2835.376	-326.350	99.799			_
29	29	3.4771	1.694	6.031	3.441	83.962	155.358	-4949.061	-315.982	98.052			_
30	30	3.8750	1.713	6.054	3.473	89.309	162.623	-9295.451	-296.468	97.150			_
31	_31	3.8750	1.618	5.974	3.308	92.912	166.505	-11290.109	-284.652	95.312			_
32											-		_
33							Ļ	ļ					
34				-									_
35				-									_
36				-									
	1												
Initial Tal	ole N	lew Table Sheet2 Sheet3									Sum	=0	

Pagina 28

Now you have the dependent variable fixed

Notice that the 8 molecular descriptors have very different magnitudes. To guarantee the comparability of the MLR coefficients (see previous lesson on QSAR), proceed to the <u>normalization</u> of all the descriptors (for instance for the interval [0; 1] ).

The first step is to calculate the MAX and MIN values of each descriptor with the functions "=MAX(range)" and "=MIN(range)" as described in the following slides

QSAR

Pagina 29

Once we have checked the dependent variable (RS) we need to analyse the independent variables (descriptors).

🏹 *Raje	sh.T	eacher.gnumeric - Gnume	ric										_ 0	×
Ele Ed	it <u>V</u>	jew Insert Format Tools	s <u>Statistic</u>	s <u>D</u> ata	Help									
C B		4 4   % 0		•	¢ -	·   🎱	$\mathbb{Z}$ fix	s, z, ≌z ¥s   m	100%					
Sans		-	10 -	A			) 🗍 🖻	] = 22	• % •	\$00 m	:   💷 (	🗐   🖽 👻		•
C33	_	+ 学 🐰 🍛	≕ =max	(C3:C31										
	Α	В	С	D	E	F	G	Н	I	J	к	L	м	
1		Dependent Variable	е		Р	hysicocl	hemical	Descriptor	5					-
2	#	RS	EA	IP	W	MR	SASA	TE	DH	SE				
3	2	0.3010	0.557	5.592	1.878	71.825	135.699	-1682.344	-440.496	117.801				
4	3	0.6020	0.502	5.619	1.830	77.927	152.919	-2451.019	-363.996	114.573				_
5	4	0.6989	0.421	5.596	1.749	71.825	137.370	-1682.335	-441.138	121.853				
6	5	0.6989	0.688	5.586	2.009	74.876	140.247	-2066.684	-400.947	116.732				_
7	6	1.3010	0.508	5.723	1.861	71.825	135.589	-1682.336	-438.212	122.617				_
8	7	1.3010	-0.175	5.614	1.277	71.825	133.584	-1682.342	-438.670	126.687				-
9	8	1.3979	0.329	5.871	1.734	77.927	152.595	-2451.015	-359.039	114.619				
10	9	1.4771	0.736	5.888	2.129	74.876	142.065	-2066.680	-401.520	114.594				-
11	10	1.6020	-0.321	5.976	1.269	63.698	131.214	-1370.073	-479.375	111.941				-88
12	11	1.6989	0.337	5.728	1.706	74.876	139.270	-2066.688	-402.632	117.804				-
15	12	1.8808	0.417	5.886	1.816	74.876	139.818	-2066.690	-399.686	118.920				-88
14	13	2.0000	0.865	5.891	2.270	77.927	146.285	-2451.029	-362.811	108.918				-
15	14	2.0791	0.559	5.790	1.926	/4.8/0	140.650	-2066.681	-402.218	101.000				-
17	15	2.0791	1.450	5.854	2.997	102.919	185.215	-21851.257	-269.022	01.010				-88
19	10	2.1700	0.455	0.018	1.881	77.027	145.549	-23/5.800	-323.333	91.919				-
19	10	2.2041	0.792	15.855	2.177	00.070	145.485	-2451.051	-224 000	110 026				-
20	10	2.3010	10.397	5.969	2.011	77.027	147 266	-2451.027	-265 520	105 421				- 1
21	20	2.3424	10.297	5 022	1 709	92 679	152 621	-2400.229	-259 419	107 722				-
22	21	2.47/1	1 286	6 000	2 817	80.011	148 470	-4564 716	-353 506	108 281				-
23	22	2.5740	0.201	5 024	1 802	76 292	146 250	-2375 798	-323 364	107 952				
24	24	2.0020	0.494	6 020	1 010	77 927	145.051	-2451 034	-363 678	100 700				-
25	25	2.9030	1 110	5 973	2 580	83 894	149 738	-6678 402	-341 980	108 112				-11
26	26	2,9030	1 250	5 920	2 751	86.878	156 431	-8792.080	-331 894	106 298				- 11
27	27	3 0000	0.543	6 115	1 989	76 235	143 952	-2475.061	-364 739	99 798				-
28	28	3.3424	0.912	6.105	2.370	80.978	151.349	-2835.376	-326.350	99,799				
29	29	3.4771	1.694	6.031	3.441	83,962	155.358	-4949.061	-315.982	98.052				
30	30	3.8750	1.713	6.054	3,473	89,309	162.623	-9295.451	-296,468	97.150				
31	31	3.8750	1.618	5.974	3.308	92.912	166.505	-11290.109	-284.652	95.312				_
33		max	=max(	3:031	IΤ									-
34		min			u.									- 22
35	1													- 11
36	1													
27														
Tottial Tab		our Table Shaata Shaata		_	_						0		- [	-

QSAR

Pagina 30

#### First we calculate the maximum and the minimum values for each colum

🐻 * Raj	esh.T	eacher.gnumeric - Gnum	ieric										
<u>File</u> Ec	lit <u>V</u>	iew Insert Format Too	ols <u>S</u> tatistics	Data	Help								
		8 8   % 6	0 🚺   🤞	-	÷		🖺 fixi	8. 2. 92 96   10	100% -				
Sans		•	10 •	AA	A		) 🗍 🖽	9 - 22	₩ % ·	\$*** a			· •
C34		💊 🐰 🖑 👻	== [=min((	C3:C31									
	Α	В	С	D	E	F	G	н	- I	J	К	L	М
1		Dependent Variab	le		Р	hysicocl	hemical	Descriptor	5				<b>A</b>
2	#	RS	EA	IP	W	MR	SASA	TE	DH	SE			
3	2	0.3010	0.557	5.592	1.878	71.825	135.699	-1682.344	-440.496	117.801			
4	3	0.6020	0.502	5.619	1.830	77.927	152.919	-2451.019	-363.996	114.573			
5	4	0.6989	0.421	5.596	1.749	71.825	137.370	-1682.335	-441.138	121.853			
6	5	0.6989	0.688	5.586	2.009	74.876	140.247	-2066.684	-400.947	116.732			
7	6	1.3010	0.508	5.723	1.861	71.825	135.589	-1682.336	-438.212	122.617		_	
8	7	1.3010	-0.175	5.614	1.277	71.825	133.584	-1682.342	-438.670	126.687			
9	8	1.3979	0.329	5.871	1.734	77.927	152.595	-2451.015	-359.039	114.619			
10	9	1.4771	0.736	5.888	2.129	74.876	142.065	-2066.680	-401.520	114.594		_	
11	10	1.6020	-0.321	5.976	1.269	63.698	131.214	-13/0.0/3	-4/9.3/5	111.941			
12	11	1.6989	0.337	5.728	1.706	74.876	139.270	-2066.688	-402.632	117.804		_	
13	12	1.8808	0.417	5.886	1.816	74.876	139.818	-2066.690	-399.686	118.920			
19	15	2.0000	0.865	5.891	2.270	71.921	146.285	-2451.029	-362.811	108.918			
10	14	2.0791	0.559	5.790	1.926	/4.8/6	140.650	-2066.681	-402.218	113.527			
17	15	2.0791	1.450	5.852	2.997	76.020	185.215	-21851.257	-209.022	01.010			
18	10	2.1700	0.455	6.018 6.022	1.881	77.027	145.549	-23/5.800	-323.333	107.919			
10	10	2.2041	0.792	5.090	2.1//	00.070	145.465	2431.031	-303.013	110.026			
20	10	2.5010	0.597	5.989	2.011	77.027	147 255	-2451.027	-265 520	105 421			
21	20	2.3424	0.757	5 022	1 709	92 679	152 621	-2431.027	-259 419	107 722			
22	20	2.47/1	1 206	6 000	2 017	82.078	149 470	-4564 716	-252 506	100.722			
23	22	2.5740	0.201	5 024	1 802	76 292	146 250	-2375 708	-222 264	107.952			
24	24	2.0020	0.391	6.020	1 010	77.027	145.051	-2451.024	-262.679	100.700			
25	25	2 9030	1 110	5 973	2 580	83 894	149 739	-6678 402	-341 980	108 112			
26	26	2.9030	1 250	5 920	2.751	86.878	156 431	-8792.080	-331 894	106 298			
27	27	3 0000	0.543	6 115	1 989	76 235	143 952	-2475 061	-364 739	99 798			
28	28	3 3424	0.912	6 105	2 370	80 978	151 349	-2835 376	-326 350	99 799		-	
29	29	3 4771	1 694	6 031	3 441	83 962	155 358	-4949.061	-315982	98 052			
30	30	3 8750	1 713	6 054	3 473	89 309	162 623	-9295 451	-296 468	97 150			
31	31	3.8750	1.618	5.974	3.308	92,912	166 505	-11290.109	-284.652	95.312			
32													
33		max	1.713										
34		min	=min(C	3:C31	ſ								
35													
36													
27	Ter.												E E
Initial Tak	ile N	ew Table Sheet2 Sheet3									Sime	0	

QSAR

Pagina 31

The method is similar to that already shown. Just do the same as displayed in the slide.

78 * R	ajesh.T	eacher.gnumeric - Gnumer	ic											1×
Ele	Edit y	jew Insert Format Tools	Statistics	<u>D</u> ata	Help									
	🖹 🖥	🗕 🕹   💥 🕠	🖸   🤞	•	<u>ب</u>	2	∑ f(x)	52 <u>5</u> 5   10	100%					
Sans		•	10 👻	A	A				9% ·	• • • • • • • • • • • • • • • • • • •		E   H •		•
C33		- 🧶 🔏 👄	= -max	(C3:C31)										
	A	В	С	D	E	F	G	н	1	J	К	L	М	
1		Dependent Variable			Р	hysicocl	hemical	Descriptor	5					-
2	#	RS	EA	IP	W	MR	SASA	TE	DH	SE				
3	2	0.3010	0.557	5.592	1.878	71.825	135.699	-1682.344	-440.496	117.801				_
4	3	0.6020	0.502	5.619	1.830	77.927	152.919	-2451.019	-363.996	114.573				_
5	4	0.6989	0.421	5.596	1.749	71.825	137.370	-1682.335	-441.138	121.853				
6	5	0.6989	0.688	5.586	2.009	74.876	140.247	-2066.684	-400.947	116.732				
- '	6	1.3010	0.508	5.723	1.861	71.825	135.589	-1682.336	-438.212	122.617				-
- 0	- /	1.3010	-0.175	5.014	1.2//	71.825	153.584	-1082.342	-458.070	114 610				
10	ő	1.3373	0.329	5 999	2 120	74.876	142.065	-2066.680	-401 520	114.019				
11	10	1.6020	-0.321	5 976	1 269	63 698	131 214	-1370.073	-479 375	111 941				-
12	11	1 6989	0 337	5 728	1 706	74 876	139 270	-2066 688	-402 632	117 804				
13	12	1.8808	0.417	5.886	1.816	74.876	139.818	-2066.690	-399.686	118,920				
14	13	2,0000	0.865	5.891	2.270	77.927	146.285	-2451.029	-362.811	108,918				_
15	14	2.0791	0.559	5.790	1.926	74.876	140.650	-2066.681	-402.218	113.527				
16	15	2.0791	1.430	5.852	2.997	102.919	183.213	-21831.257	-269.022	101.888				
17	16	2.1760	0.453	6.018	1.881	76.830	145.349	-2375.806	-323.533	91.919				_
18	17	2.2041	0.792	5.833	2.177	77.927	145.483	-2451.031	-365.015	107.813				_
19	18	2.3010	0.597	5.989	2.011	80.978	157.941	-2835.362	-324.988	110.836				-81
20	19	2.3424	0.737	5.911	2.135	77.927	147.355	-2451.027	-365.530	105.431				-81
21	20	2.47/1	0.387	5.923	1.798	82.678	152.621	-2490.338	-358.418	107.722				-
22	21	2.5/40	1.286	6.009	2.817	80.911	148.479	-4564./16	-353.596	108.281				
24	24	2.0020	0.391	5.924	1.802	77.027	146.250	-2451 024	-262.679	107.833				
25	24	2.0129	1 1 1 0	5 073	2 580	93 904	140.739	-6678 402	-341 090	109.790				
26	26	2.9030	1.250	5 920	2.560	86 878	156 431	-8792.080	-331 894	106 298				
27	27	3 0000	0.543	6 115	1 989	76 235	143 952	-2475.061	-364 739	99 798				
28	28	3.3424	0.912	6.105	2.370	80.978	151.349	-2835.376	-326.350	99,799				
29	29	3.4771	1.694	6.031	3.441	83.962	155.358	-4949.061	-315.982	98.052				
30	30	3.8750	1.713	6.054	3.473	89.309	162.623	-9295.451	-296.468	97.150				
31	31	3.8750	1.618	5.974	3.308	92.912	166.505	-11290.109	-284.652	95.312				_
32														_
33		max	1.713											
34		min	0.321	F -										-81
30	_													-8
27														
	4													2
Initial 1	Table N	ew Table Sheet2 Sheet3									Sum=1.39	2		1

Pagina 32

QSAR

🐻 * Ra	ijesh.T	eacher.gnumeric - Gnumer	ic									- D ×
Ele	Edit y	jew Insert Format ⊥ools	Statistics	s <u>D</u> ata <u>H</u> elp								
	6	]   🕹 🕹   💥 🕠	🖸   🍕	, - 🧇 -	0	$\sum f(\mathbf{x})$	52 Si   10	100%				
Sans	_		10 👻	AAA				9% ·	<b>∲</b> ‰ _∞		]  🖽 🔻	•
C33		💊 🐰 🖑 👻	= -max	(C3:C31)								
	A	В	С	DE	F	G	н	1	1	К	L	м
1		Dependent Variable		Р	hysicoc	hemical	Descriptor	5				*
2	#	RS	EA	IP W	MR	SASA	TE	DH	SE			
3	2	0.3010	0.557	5.592 1.878	71.825	135.699	-1682.344	-440.496	117.801			
4	3	0.6020	0.502	5.619 1.830	77.927	152.919	-2451.019	-363.996	114.573			
5	4	0.6989	0.421	5.596 1.749	71.825	137.370	-1682.335	-441.138	121.853			
6	5	0.6989	0.688	5.586 2.009	74.876	140.247	-2066.684	-400.947	116.732			
7	6	1.3010	0.508	5.723 1.861	71.825	135.589	-1682.336	-438.212	122.617			
8	7	1.3010	-0.175	5.614 1.277	71.825	133.584	-1682.342	-438.670	126.687			
- 10	8	1.3979	0.329	5.8/1 1.734	77.927	152.595	-2451.015	-359.039	114.619			
11	9	1.4//1	0.730	5.888 2.129	/4.8/6	142.065	-2066.680	-401.520	114.594			
12	10	1.6020	-0.321	5.970 1.209	03.098	131.214	-13/0.0/3	-4/9.5/5	117.941			
13	12	1.0989	0.337	5.728 1.700	74.870	120 010	-2066.600	-200.696	112 020			
14	12	2,0000	0.965	5 801 2 270	77 027	146 295	-2451.020	-362.811	108 018			
15	14	2.0000	0.559	5 790 1 926	74 876	140.285	-2066 681	-402.218	113 527			
16	15	2 0791	1 4 3 0	5 852 2 997	102 919	183 213	-21831 257	-269 022	101 888			
17	16	2.1760	0.453	6.018 1.881	76.830	145.349	-2375.806	-323.533	91,919			
18	17	2,2041	0.792	5.833 2.177	77,927	145,483	-2451.031	-365.015	107.813			
19	18	2.3010	0.597	5.989 2.011	80.978	157.941	-2835.362	-324.988	110.836			
20	19	2.3424	0.737	5.911 2.135	77.927	147.355	-2451.027	-365.530	105.431			
21	20	2.4771	0.387	5.923 1.798	82.678	152.621	-2490.338	-358.418	107.722			
22	21	2.5740	1.286	6.009 2.817	80.911	148.479	-4564.716	-353.596	108.281			
23	22	2.6020	0.391	5.924 1.802	76.383	146.250	-2375.798	-323.364	107.853			
24	24	2.8129	0.494	6.020 1.919	77.927	145.051	-2451.034	-363.678	109.790			
25	25	2.9030	1.110	5.973 2.580	83.894	149.738	-6678.402	-341.980	108.112			
20	26	2.9030	1.250	5.920 2.751	86.878	156.431	-8792.080	-331.894	106.298			
2/	2/	3.0000	0.543	6.115 1.989	76.235	143.952	-24/5.061	-364.739	99.798			
20	20	2.4771	1.604	6.021 2.441	80.978	151.549	-4040.061	-215 092	09.799			
30	29	3.4771	1.094	6.051 3.441	83.962	162,622	-4949.061	-206.469	98.052			
31	31	3.8750	1.618	5 974 3 308	92 912	166 505	-11290 100	-284 652	95 312			
32	- 51	5.8750	1.018	5.574 5.508	72.912	100.505	11230.109	204.032	25.512			
33		max	1.713	max(J3:J31)	*****			(******************	_ 1			
34		min	0.321	min(J3:J31)								
35				g	************	*******	*************************					
36												
27	1											
Initial T	able N	lew Table Sheet2 Sheet3			Drag to	o autofil				Sum=1.393	2	

Pagina 33

QSAR

🏹 *Raje	sh.T	eacher.gnumeric - Gnumeri	ic											١×
Ele Ed	it <u>V</u>	jew Insert Format Tools	Statistic	s <u>D</u> ata	Help									
D E		🗕 🖴   🔏 🕞	🖸   🍕	•	() ·	9	$\sum f(\mathbf{x})$	Sz 🖏   🛍	100%	•				
Sans			10 👻	A	A				9%	. 🎭 👌		1 🖽	•	•
C33	_	🥥 🐰 🖑 🔻 :	= -max	(C3:C31	)									
	Α	В	С	D	E	F	G	н	I	J	К	L	м	
1		Dependent Variable			Р	hysicocl	hemical	Descriptors	3					-
2	#	RS	EA	IP	W	MR	SASA	TE	DH	SE				
3	2	0.3010	0.557	5.592	1.878	71.825	135.699	-1682.344	-440.496	117.801				
4	3	0.6020	0.502	5.619	1.830	77.927	152.919	-2451.019	-363.996	114.573				
5	4	0.6989	0.421	5.596	1.749	71.825	137.370	-1682.335	-441.138	121.853				
6	5	0.6989	0.688	5.586	2.009	74.876	140.247	-2066.684	-400.947	116.732				
/	6	1.3010	0.508	5.723	1.861	71.825	135.589	-1682.336	-438.212	122.617				
8	7	1.3010	-0.175	5.614	1.277	71.825	133.584	-1682.342	-438.670	126.687				
9	8	1.3979	0.329	5.871	1.734	77.927	152.595	-2451.015	-359.039	114.619				
10	9	1.4//1	0.736	5.888	2.129	74.876	142.065	-2066.680	-401.520	114.594				
12	10	1.6020	-0.321	5.976	1.269	63.698	131.214	-13/0.0/3	-4/9.3/5	111.941				
12	11	1.6989	0.337	5.728	1.706	74.876	139.270	-2066.688	-402.632	117.804				
14	12	1.8808	0.417	5.880	1.810	74.870	139.818	-2066.690	-399.080	108.920				
15	15	2.0000	0.805	5.891	2.270	74.927	140.285	-2451.029	-302.811	108.918				
16	14	2.0791	0.559	5.790	1.920	/4.8/0	140.050	-2000.081	-402.218	101.000				
17	16	2.0791	0.452	5.852	1 001	76 920	145 240	-2275 906	-209.022	01.010				
18	17	2.1700	0.455	5 022	2 177	77.027	145.349	-2451.021	-265 015	107 912				
19	10	2.2041	0.792	5 090	2.177	20.079	145.465	-2431.031	-224 000	110 926				
20	10	2.3010	0.337	5 011	2.011	77 027	147 255	-2451.027	-265 520	105 421				
21	20	2.3424	0.397	5 023	1 709	82.678	152 621	-2400.338	-359 419	107 722				
22	21	2.5740	1 286	6 000	2 917	80.011	149 470	-4564 716	-252 506	109.201				
23	22	2.5740	0.301	5 924	1 802	76 383	146 250	-2375 798	-323 364	107.853				
24	24	2.0020	0.101	6 020	1 010	77 027	145.051	-2451.034	-363.679	100 700				
25	25	2 9030	1 110	5 973	2 580	83 894	149 738	-6678 402	-341 980	108 112				
26	26	2.9030	1 250	5 920	2 751	86 878	156 431	-8792.080	-331 894	106 298				
27	27	3 0000	0.543	6 115	1 989	76 235	143 952	-2475.061	-364 739	99 798				
28	28	3 3424	0.912	6 105	2 370	80 978	151 349	-2835 376	-326350	99 799				
29	29	3 4771	1 6 9 4	6 031	3 441	83 962	155 358	-4949 061	-315,982	98 052				
30	30	3.8750	1.713	6.054	3.473	89,309	162.623	-9295,451	-296,468	97.150				
31	31	3.8750	1.618	5.974	3.308	92.912	166.505	-11290.109	-284.652	95.312				
32		max	1.713	6.115	3.473	102.919	183.213	1370.073	269.022	126.687				
34		min	0.321	5.586	1.269	63.698	131.214	21831.257	479.375	91.919				
35														
36						N								
27	1													<u>ک</u>
Initial Tab	le N	ew Table Sheet2 Sheet3							_		Sum=23222	742		Ξ.

QSAR

Pagina 34

#### And all the min and max values are promptly calculated

*Rajesh.	leacher.gnumeric - (	Gnumeric	atistics Da	ta Helo								<u>_                                    </u>
						f(x) = = = =	žs I ša I	100%				
						)~~ <u>¥z</u>	24 n   2005		A			
ans		▼ 10	• A	AA		-		9 % ·	<b>*</b> ***********************************		🖽 🔻	•
1	🍛 🐰 🍳	· - =										
	A B	С	D	E	F	G	н	1	J	К	L	м
1												*
2												
4												
5												
6												
8												
9												
10												
11												
13												
14												
15												
16												
1/												
19												
20												
21												
22												
23												
25												
26												
27												
28												
30												
31												
32												
33												
35												
36												
37												
al Table It	New Table Sheet2 Sh	elita								Sum=0		<u> </u>
	Jane Jane Jane	1. A. A.								Jumeo		1

QSAR

Pagina 35

In a new sheet ....

Sans			▼ 10	- A	AA			- 88   4	9 % ·	۰. ۵			•	
A1		o 🕺 🔌	· =											
_	Α	в	с	D	E	F	G	н	1	J	к	L	M	
1													*	
2														
4														
5														
6														
7														
8														
10									-					
11														
12														
13														
14														
15														
17														
18														
19														
20														
21														
23														
24														
25														
26														
28						-								
29														
30														
31														
32														
34														
35														
36													-	
37	1													
	New Table	Sheet2 No	ormalized								Sum=0			
Initial Table														

... renamed «Normalized» ...
				5 <u>0</u> 010	diap -		ST 6(1)	s z. Láo	1008/				
				<b>∛</b> *	· ·	<b></b>	21 JW :	Miz Min   III	100%				
s Ne	w Roman	-	12 🔻	<b>A</b>	4 <u>A</u>		) 🗉 🎛		B % ·	• 🎭 🔤	: 🗐 🖬		•
_		💊 🐰 🎺 👻	= #										
	A	В	С	D	E	F	G	н	1		к	L	M
1	Dep	endent Variab	e		P	hysicocl	nemical	Descriptors	5			,	
2	#	N. Ort		IP	W	MR	SASA	TE	DH	SE			
3	2	ee cui	7	5.592	1.878	71.825	135.699	-1682.344	-440.496	117.801			
ŧ	3	Сору	2	5.619	1.830	77.927	152.919	-2451.019	-363.996	114.573			
5	4	Paste	1	5.596	1.749	71.825	137.370	-1682.335	-441.138	121.853			
5	5	Paste Special	8	5.586	2.009	74.876	140.247	-2066.684	-400.947	116.732			
/	6	Insert 66 Cells	8	5.723	1.861	71.825	135.589	-1682.336	-438.212	122.617			
3	7	Delete 66 Cell	75	5.614	1.277	71.825	133.584	-1682.342	-438.670	126.687			
)	8	A Clean Contract		5.871	1.734	77.927	152.595	-2451.015	-359.039	114.619			
0	9	Clear Content	s 6	5.888	2.129	74.876	142.065	-2066.680	-401.520	114.594			_
1	10	Add Comment	21	5.976	1.269	63.698	131.214	-1370.073	-479.375	111.941			
2	11		/	5.728	1.706	74.876	139.270	-2066.688	-402.632	117.804			
3	12	Add Hypenink	/	5.886	1.816	74.876	139.818	-2066.690	-399.686	118.920			
5	13	Eormat 66 Cel	s D	5.891	1.026	74.927	140.285	-2451.029	-302.811	108.918			
6	14	Cel	• •	5.790	2.007	102.010	102 212	-21021 257	-260.022	101 000			-
7	16	Column	• 2	6.019	1 991	76 830	145 340	-2375 806	-209.022	01 010			
8	17	Row	• 6	5 833	2 177	77 927	145 483	-2451.031	-365.015	107 813			-
9	18	2 3010	0.597	5 989	2 011	80.978	157 941	-2835 362	-324 988	110.836			
0	19	2 3424	0.737	5 911	2 135	77 927	147 355	-2451.027	-365 530	105 431			
1	20	2,4771	0.387	5.923	1.798	82.678	152.621	-2490.338	-358.418	107,722			
2	21	2.5740	1.286	6.009	2.817	80.911	148.479	-4564.716	-353.596	108.281			
3	22	2,6020	0.391	5.924	1.802	76.383	146.250	-2375.798	-323.364	107.853			
4	24	2.8129	0.494	6.020	1.919	77.927	145.051	-2451.034	-363.678	109.790			
5	25	2.9030	1.110	5.973	2.580	83.894	149.738	-6678.402	-341.980	108.112			
6	26	2.9030	1.250	5.920	2.751	86.878	156.431	-8792.080	-331.894	106.298			
7	27	3.0000	0.543	6.115	1.989	76.235	143.952	-2475.061	-364.739	99.798			
8	28	3.3424	0.912	6.105	2.370	80.978	151.349	-2835.376	-326.350	99.799			-
9	29	3.4771	1.694	6.031	3.441	83.962	155.358	-4949.061	-315.982	98.052			-
0	50	3.8750	1.713	6.054	3.473	89.309	162.623	-9295.451	-296.468	97.150			
1	51	3.8750	1.618	5.974	3.308	92.912	166.505	-11290.109	-284.652	95.312			
2			1 71 2	6 115	2 4 7 2	102.010	102 212	1270 072	260.022	126 607			-
4		min	0.221	5 5 9 6	3.473	62 609	121 214	21021 257	470 275	01 010			
5		11001	0.521	3.380	1.209	05.098	151.214	21051.237	479.575	91.919			-
6													+
-													+

QSAR

Pagina 37

#### Copy the ID and labels from «New Table» and also the descriptors labels



... paste everything in the «Normalized» sheet

Sans		•	10	• [	A	4 <u>A</u>					- 88	<b>19</b> 2	ι.	<b>4</b>		- 68 -		-	
L21	۵ 😵	<u>ن</u> ب	- [																
	A B		С	DE	F	G	Н	1	J	К	L		м	N	0	Р	Q	_	
1	# RS		EA	IP W	/ MR	SASA	A TE	DH	SE									<u>^</u>	
2	2 0.301	0																	
3	3 0.602	0					5	2											
4	4 0.698	9																	
5	5 0.698	9																	
6	6 1.301	0																	
7	7 1.301	0																	
8	8 1.397	9																	
9	9 1.477	1																	
10	1.602	20														-			
11	1 1.698	39																	
12	1.880	8		_							_						_		
13	13 2.000	00		_	-		_											-	
14	14 2.079	21		_	-		_												
15	5 2.079	1		_	-		-				_							-	
10	2.176	0		-	-		-				_						-	-	
10	2.204	1		-	-		-												
10	2.301	0		-	-		-				-						-	-	
20	2.542	4		-	-		-										-	-	
21	20 2.477	1		-	-														
22	2.57	0									<u> </u>	-							
23	2 2.002	0																	
24	2 2 903	0																	
25	26 2.903	0																	
26	27 3 000	00																	
27	3.342	4																	
28	3.477	1																	
29	30 3.875	50																	
30	3.875	50																	
31																			
32	max	<		_														-	
33	min			_	-	-	-									-			
34				_	-	-	-				_					-	-		
35				_	-	-	-				_								
50																		- <b>-</b>	
	<u>•</u>																Þ	1	
Initial Table	New Table Sheet2	Normalized	J												Sum=0			11.	
								-											

Make sure you have the same values for RS

Sans		•	10	•	A	A	A				89	8	=   %	9 %	. (	00 m00			• •	•	•
C2		😜 🐰 🛫 👻	-	=(Nev	w Table	e'!C3	'New T	able'!	C\$34)/	(New	Table'	C\$33-'N	ew Table	'IC\$34	)						
	Α	В	0		DE	F	G	H	I I	J	ł	<	L		м	N		0	Р	(	
1	#	RS	E	A I	PW	MR	SAS	AT	E DH	I SE											A
2	2	0.3010	=("	New	Table	e'!C	3-'Nev	v Tal	ble'!C	\$34	)/('Ne	🖗 Tabl	e'!C\$3	3-'Ne	w Table	e'!C\$34	)				
3	3	0.6020		- T								~									
4	4	0.6989						_		_											
5	5	0.6989		_	_			_	_	_				_							
6	6	1.3010	_	_	_			_	-	-				_							
	1	1.3010	_	-	-			-		-				_							
8	8	1.3979	_	-	-			-		-				_							
10	9	1.4//1	_	-				-	-	-											
11	11	1.6020	_	-	-			-	-	-				-							
12	12	1.8808		-				-	-	-											
13	13	2.0000																			
14	14	2.0791																			
15	15	2.0791																			
16	16	2.1760																			
17	17	2.2041																			
18	18	2.3010																			
19	19	2.3424			_			_		_											
20	_20	2.4771	_	_	_			_	-	_				_						_	
21	21	2.5740	_	_	_			-	-	-				_						_	
22	22	2.6020	_	-				-	-	-				-						_	
23	24	2.8129		-				-	-	-				-							
24	25	2.9030		-	-			-	-	-				-							
26	27	3,0000		-						-											
27	28	3 3424																			
28	29	3.4771																			
29	30	3,8750																			
30	31	3.8750																			
31																					
32		max		_	_			_		_										_	
33		min	_	_				_	_					_						_	
34			_	_	_			_	_	_				_						_	
30			_	-				-	-	-				_						_	
30																					<b>*</b>
	1	and the second second																		<u>}</u>	
Initial Ta	ple New	Table Sheet2 Normali	zed														Sum=0.4	132			11.

Now apply the normalization as described in the lesson. (see also the formula in the slide). Here the MinMax normalization is being applied: Normalized valued = (original value – min) / (max –min)

	8 🗖   🖴	4	¥ (	•		- «	» -	0	$\sum f(x)$	1. z. 1. ž	s   1	<b>A</b> 100	)%	•				
s			_	▼ 10	- (	A A	A			<b>PP</b>	88	ų,	36	. 🧠		-	E8 -	•
		<u>s</u> %	18 .	-	-may(C	2-030)												
	AB			- 1	E	L.(5,0)	ы			ĸ	1	1		4	N	0	D	
1	# RS	ΕA	IP	w	MR	SASA	TE	DH	SE	IX.	_	-						
2	2 0 3010	0.422	0.011	0.276	0.007	0.000	0.005	0.105	0.744									
2	2 0.5010	0.452	0.011	0.276	0.207	0.080	0.963	0.103	0.744		_							
1	3 0.0020	0.405	0.002	0.235	0.303	0.110	0.947	0.343	0.052		-							
5	5 0 6080	0.303	0.019	0.216	0.207	0.174	0.903	0.102	0.001									
6	6 1 3010	0.408	0.259	0.269	0.207	0.084	0.985	0.196	0.883									
7	7 1 3010	0.072	0.053	0.004	0.207	0.046	0.985	0.194	1.000									
8	8 1 3070	0.320	0.539	0 211	0.363	0.411	0.947	0.572	0.653									
	9 1 4771	0.520	0.571	0.390	0.285	0.209	0.966	0.370	0.652									
0	10 1 6020	0.000	0.737	0.000	0.000	0.000	1.000	0.000	0.576									
1	11 1 6989	0.324	0.268	0.198	0.285	0.155	0.966	0.365	0.745									
2	12 1 8808	0.363	0.567	0.248	0.285	0.165	0.966	0.379	0.777									
3	13 2,0000	0.583	0.577	0.454	0.363	0.290	0.947	0.554	0.489									
4	14 2.0791	0.433	0.386	0.298	0.285	0.181	0.966	0.367	0.621									
.5	15 2.0791	0.861	0.503	0.784	1.000	1.000	0.000	1.000	0.287									
.6	16 2.1760	0.381	0.817	0.278	0.335	0.272	0.951	0.741	0.000									
.7	17 2.2041	0.547	0.467	0.412	0.363	0.274	0.947	0.544	0.457									
8	18 2.3010	0.451	0.762	0.337	0.441	0.514	0.928	0.734	0.544									
9	19 2.3424	0.520	0.614	0.393	0.363	0.310	0.947	0.541	0.389									
0	20 2.4771	0.348	0.637	0.240	0.484	0.412	0.945	0.575	0.455									
1	21 2.5740	0.790	0.800	0.702	0.439	0.332	0.844	0.598	0.471									
2	22 2.6020	0.350	0.639	0.242	0.323	0.289	0.951	0.742	0.458									
3	24 2.8129	0.401	0.820	0.295	0.363	0.266	0.947	0.550	0.514									
4	25 2.9030	0.704	0.732	0.595	0.515	0.356	0.741	0.653	0.466									
5	26 2.9030	0.772	0.631	0.672	0.591	0.485	0.637	0.701	0.414									
6	27 3.0000	0.425	1.000	0.327	0.320	0.245	0.946	0.545	0.227									
/	28 3.3424	0.606	0.981	0.500	0.441	0.387	0.928	0.727	0.227									
8	29 3.4771	0.991	0.841	0.985	0.517	0.464	0.825	0.777	0.176		_							
9	30 3.8750	1.000	0.885	1.000	0.653	0.604	0.613	0.870	0.150		-						_	
0	31 3.8750	0.953	0.733	0.925	0.745	0.679	0.515	0.926	0.098		-					-	_	
1											-							
<b>Z</b>	max	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000		-							
3	min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		-							
4											-							
0											-					-		

QSAR

Pagina 41

And let's obtain the normalized matrix, that can be checked by calculating min and max. Their values should range all between 0 and 1.



At this point we can continue analysing the data.

132 1 2 3 4 5 W 6 0 1 1 1 2 R 3 E 4 1 1 2 R 5 V 6 7 7 7 7 7 7 7 7 7 7 7 7 7	A B RS S	C EA	D	E	F								
1 2 R 3 E 4 I 5 W 6 M	A B RS S	C EA	D	E	F								
1 2 R 3 E 4 I 5 V 6 M	RS A	EA	10			G	H			ĸ	L	M	
2 R 3 E 4 I 5 V 6 M	A A		Ir	W	MR	SASA	TE	DH	SE			<u>^</u>	1
4 II 5 V 6 M	W1												
5 V 6 M	P												
6 M	w		-	-	-			-					
7 64	/IR		¢										
/ 5A	SA												
8 T	Æ												
9 D	H												
10 S	SE												
11													
12													
14		-	-				-		-				
15													
16													
18													
19													
20													
21										++			
23				+						+			
24													
25													
26										++			
28													
29													
30													
31			+		+			<u> </u>	1				
33				-	-				•				
34													
35													4
						-		F					
Initial Table Y ch	heck New Table	Normalized	Intercorrelation	n Sheet3						Sum=0			1

Prepare a new sheet «Intercorrelation» as in the slide

Sans			• 10	- A	AA			88 4	9 % ·	\$00 800 \$00 800			-	
82		🧧 🗶 🛫	• ==	=rsq										
	Α	В	С	[rsq :	s	quare of the	Pearson corre	lation coeffici	ient of the pa	ired set of dat	a] K	L	M	
1		RS	EA	IP	W	MR	SASA	TE	DH	SE			*	
2	RS	=rsq[]												
3	EA													
4	IP													
5	W													
6	MR													
7	SASA													
8	TE													
9	DH													
10	SE													
11														
12														
13														
15				-										
16				-										
17														
18														
19														
20														
22														
23														
24														
25														
26														
27				-										
29														
30														
31														
32														
33														
35				-									-1	
55	4													
Initial Tabl	e Y check I	New Table N	ormalized 1	intercorrelatio	n Sheet3						Sum=0		4	
_													10	

And by the proper function calculate the squared pearson coefficients (r2)

		ے 🐰 💫	· =	=rsq(Normali	zed!B2:B30,N	Normalized !B2	::B30)				
	Α	В	С	D	E	F	G	н	1	J	
1		RS	EA	IP	W	MR	SASA	TE	DH	SE	
2	RS	1									
3	EA										
4	IP										
5	W										
7	NIK										
/	SASA										
0											
10	SE										
11	5E										Ţ
		!				1	1	1		Þ	ſ
Y check	New Table	Normalized	Intercorrel	ation Sheet	3 •			Sum=1			/
9 10 11 Y check	DH SE	Normalized	Intercorrel	ation Sheet	3 1			Sum=1			•

First RS vs RS. It should be = 1!



Lock at the formula and define fixed variable (insert \$ signs as in the slide)



#### Then just by drugging the r2 is calculated for all the values in the first row. Make a similar operation in the others rows

🐻 *Raje	sh.Teacl	her.gnu	meric -	Gnume	ric							IX			
<u>Eile E</u> di	t <u>V</u> iew	<u>I</u> nsert	F <u>o</u> rma	t <u>T</u> ools	<u>S</u> tatis	tics <u>D</u> a	ata <u>H</u> elp	)							
		8	3   2	6		⇔ •	· 🚸	<b>v</b>	🕥 Σ	f(x)	n. Mz	•			
Sans				•	10 -	A	AA				3 🖽	•			
19	I9         Image: Second s														
	Α	В	С	D	E	F	G	Н	I		к				
1		RS	EA	IP	W	MR	SASA	TE	DH	ŚÈ		-			
2	RS	1.000	0.390	0.728	0.469	0.340	0.282	0.156	0.517	0.614					
3	EA		1.000	0.146	0.975	0.649	0.532	0.454	0.550	0.388					
4	IP			1.000	0.201	0.129	0.150	0.043	0.339	0.597					
5	W				1.000	0.648	0.543	0.479	0.564	0.436					
6	MR					1.000	0.953	0.496	0.776	0.085					
7	SASA						1.000	0.117	0.798	0.025					
8	TE							1.000	0.023	0.290					
9	DH								1.000	0.058					
10	SE									1.000					
11	1											⊢⊢			
Normal	ized	Intercor	relation	•				Sum=1	.058						

QSAR

Pagina 48

#### And the correlation matrix is obtained

🐻 * Raje	esh.Teacl	her.gnu	meric -	Gnume	ric				Ν		_ 🗆	×
<u>Eile E</u> o	lit <u>V</u> iew	<u>I</u> nsert	Format	t <u>T</u> ools	<u>S</u> tatis	tics <u>D</u> a	ata <u>H</u> elp	)	N			
	3 🖬	8	3   🔉			→	· 🚸	-	<b>e</b>	f(x)	a. Mz	•
Sans				•	10 -	A	AA			1111	] 🗆	•
СЗ		٩	∦ ≪	-	= [=rs	sq(Norma	lized!C2:	C30,Nor	malized!	C2:\$C30	))	
	Α	В	С	D	E	F	G	Н	I	J	К	
1		RS	EA	ID	W	MR	SASA	TE	DH	SE		*
2	RS	1.000	0.390	0.728	0469	0.340	0.282	0.156	0.517	0.614		
3	EA		1.000	0.140	0.975	0.649	0.532	0.454	0.550	0.388		
4	IP			1.000	0.201	0.129	0.150	0.043	0.339	0.597		
5	W				1.000	0.648	0.543	0.479	0.564	0.436		
6	MR					1.000	0.953	0.496	0.776	0.085		
7	SASA						1.000	0.117	0.798	0.025		
8	TE							1.000	0.023	0.290		
9	DH								1.000	0.058		
10	SE									1.000		
11	1										1	
• Norma	alized	Intercor	relation	•				Sum=19	.445			

QSAR

Pagina 49

In the first row are reported the r2 values for each variable with the dependent variable, these are monoparametric regressions. The best one is IP that as the highest r^2 value.

🐻 * Raje	sh.Teacl	ier.gnu	meric -	Gnume	ric						<u>_   ×   </u>
<u>F</u> ile <u>E</u> di	t <u>V</u> iew	<u>I</u> nsert	F <u>o</u> rma	t <u>T</u> ools	<u>S</u> tatist	ics <u>D</u> ata	a <u>H</u> elp				
C 🗅		8	l   😹	6		<del>⇔</del> →		6	) 🛛 .	f(x) 🔒 z	-
Sans				•	10 -	A	AA				
B11		٩	∦ ∢	ļ -	=						
	Α	В	С	D	Е	F	G	н	1	J	к
1		RS	EA	IP	W	MR	SASA	TE	DH	SE	-
2	RS	1.000	0.390	0.728	0.469	0.340	0.282	0.156	0.517	0.614	
3	EA		1.000	0.146	0.975	0.649	0.532	0.454	0.550	0.388	
4	IP			1.000	0.201	0.129	0.150	0.043	0.339	0.597	
5	W				1.000	0.648	0.543	0.479	0.564	0.436	
6	MR					1.000	0.953	0.496	0.776	0.085	
7	SASA						1.000	0.117	0.798	0.025	
8	TE							1.000	0.023	0.290	
9	DH		<del>c</del>						1.000	0.058	
10	SE									1.000	
			ľ								
• Normal	ized	Intercor	relation	•				Sum=0			

QSAR

Pagina 50

In the yellow area are highlighted the intercorrelations among the descriptors to check for any collinearity. As a limit an r2 greater than 0.5 means that two variable are correlated and should not used together.

In order to get an overview of the uniparametric relationship between the various descriptors we use a linear regression analysis function built into Gnumeric

(Statistics > Dependent Observations > Regression)

QSAR

Pagina 51

In the next slides we are going to make a stepwise building of the QSAR model

🧒 * R	lajes	h.Teacl	her.gnu	meric -	Gnumeri	c										_	
Eile	<u>E</u> dit	<u>V</u> iew	Insert	Forma	t <u>T</u> ools	Statistic	s <u>D</u> ata	<u>H</u> elp									
			8	3   2		<u>D</u> es San	criptive S	tatistics		-	f(x) 🕺 z	2. Ma   111	100%	-			
Sans					<b>•</b> 1	Dep	endent O	bservatio	ns		F <u>o</u> reca Fourier	st		• =	8 += +=		-
B11			٩	* «		<u>O</u> ne <u>T</u> we	e Sample 1 o Sample 1	Fests Fests		<b>*</b>	<u>K</u> aplan	-Meier Estima	ites				
		Α	В	С	D	Mul	tiple Samp	le Tests		<u> </u>	Reares	sion	ta Andryalam	·	N	0	
1			RS	EA	IP	W	MR	SASA	TE	DH	SE		N T				-
2		RS	1.000	0.390	0.728	0.469	0.340	0.282	0.156	0.51	7 0.614						
3		EA		1.000	0.146	0.975	0.649	0.532	0.454	0.550	0.388						
4		IP			1.000	0.201	0.129	0.150	0.043	0.33	9 0.597						
5		W				1.000	0.648	0.543	0.479	0.564	4 0.436						
6		MR					1.000	0.953	0.496	0.77	<mark>6</mark> 0.085						
7		SASA						1.000	0.117	0.798	B 0.025						
8		TE							1.000	0.02	3 0.290						
9		DH								1.00	0.058						
10	)	SE									1.000						
11	L																
12	2																
14	2																+
15	5																
16	5																+
17	7																
		4															
Initial	Table	Y chec	k New 1	Table N	ormalized	Interco	relation	Sheet3			Regres	sion Analysis	5	Sur	n=0		

QSAR

Bit giver insert Fyrmat Tools       Statistics       Data Help         Descriptive Statistics       Descriptive Statistics       Image: Construction of the statistics of the statis of the statistics of the statistics of the s	🐻 * Raje	sh.Teac	her.gnu	meric -	Gnumer	ic				
Image: Construction of the statistics in the statistex in the statistics in the statistics in the statistic	<u>File E</u> d	it <u>V</u> iew	<u>I</u> nsert	F <u>o</u> rma	t <u>T</u> ools	Statistic	s <u>D</u> ata	<u>H</u> elp		
Sans       I       Dependent Observator       Imput Options Qutput       Imput Options Qutput         311       Imput Options Qutput       Imput Options Qutput       Imput Options Qutput       Imput Options Qutput         1       RS       EA       IP       W       MR       SASA         2       RS       1.000       0.390       0.7228       0.469       0.340       0.282 (Interview)       Imput Options Qutput       Imput Options Qutp	C 🖻		8	3   8	6	Des San	scriptive S npling	tatistics	f(x) 👷 Z 🙀 👔 100% 🗸	
311       Ope Sample Tests Invo Sample Tests Multiple Sample Tests       Input Options Qutput         1       RS       EA       IP       W MR       SASA         2       RS       1.000       0.390       0.728       0.469       0.340       0.282         3       EA       1.000       0.390       0.728       0.469       0.532       Multiple linear regression       N       O         3       EA       1.000       0.201       0.129       0.150       Multiple dependent (y) variables       Imput Options       Multiple dependent (y) variables       Imput Options       Imput O	Sans				-	Dep	pendent (	bservation	Regression	-
A       B       C       D       Multiple Sample Tests         1       RS       EA       IP       W       MR       SASA         2       RS       1.000       0.390       0.728       0.469       0.340       0.282 (         3       EA       1.000       0.216       0.975       0.649       0.532 (       Multiple 2-yariable regressions       Imultiple 2-yariable regressions         5       W       Imultiple 0.201       0.129       0.150 (       Imultiple 0.543 (       Imultiple 0	B11			* «		<u>O</u> ne <u>T</u> we	e Sample o Sample	Tests Tests	Input Options Qutput	
1       RS       EA       IP       W       MR       SASA         2       RS       1.000       0.390       0.728       0.469       0.340       0.282         3       EA       1.000       0.146       0.975       0.649       0.532       Multiple 2-yariable regressions         4       IP       1.000       0.201       0.129       0.150       Multiple dependent (y) variables         5       W       1.000       0.648       0.543       Multiple dependent (y) variables         6       MR       1.000       0.953       Multiple dependent (y) variables         9       DH       Variables:       Normalized! \$C\$1:\$C\$30       Yariables:         10       SE       Variables:       Normalized! \$C\$1:\$C\$30       Yariable:         12       I       I       I       I       I       I         13       I       I       I       I       I       I       I         15       I <t< td=""><td></td><td>A</td><td>В</td><td>С</td><td>D</td><td>Mul</td><td>tiple Sam</td><td>ole Tests</td><td>N</td><td>0</td></t<>		A	В	С	D	Mul	tiple Sam	ole Tests	N	0
2       RS       1.000       0.390       0.728       0.469       0.340       0.282       C         3       EA       1.000       0.146       0.975       0.649       0.532       C       Multiple 2-yaniable regressions       C       Multiple dependent (y) variables         5       W       1.000       0.648       0.543       C       Multiple dependent (y) variables       C       C         6       MR       1.000       0.953       C       Multiple dependent (y) variables       C	1		RS	EA	IP	W	MR	SASA	<u>M</u> ultiple linear regression	<b>A</b>
3       EA       1.000       0.146       0.975       0.649       0.532       ()         4       IP       1.000       0.201       0.129       0.150       ()       ()       variables       ()       ()       variables       ()	2	RS	1.000	0.390	0.728	0.469	0.340	0.282 0	Multiple 2-variable regressions	
4       IP       1.000       0.201       0.129       0.150         5       W       1.000       0.648       0.543         6       MR       1.000       0.953         7       SASA       1.000       0.953         7       SASA       1.000       0.953         9       DH       1.000       1.000         8       TE       1.000       1.000         9       DH       1.000       1.000         10       SE       1.000       1.000         12       1.000       1.000       1.000         13       1.000       1.000       1.000         14       1.000       1.000       1.000         16       1.000       1.000       1.000         17       1.000       1.000       1.000         16       1.000       1.000       1.000       1.000         16       1.000       1.000       1.000       1.000         17       1.000       1.000       1.000       1.000         16       1.000       1.000       1.000       1.000         16       1.000       1.000       1.000       1.000 <td>3</td> <td>EA</td> <td></td> <td>1.000</td> <td>0.146</td> <td>0.975</td> <td>0.649</td> <td>0.532</td> <td></td> <td></td>	3	EA		1.000	0.146	0.975	0.649	0.532		
5       W       1.000 0.648 0.543 ( 1.000 0.953 ( 7 SASA       1.000 0.953 ( 1.000 ( 8 TE         9       DH       1.000 0.648 0.543 ( 1.000 0.953 ( 8 TE       1.000 ( 8 TE         9       DH       1.000 0.648 0.543 ( 1.000 0.953 ( 8 TE       1.000 ( 9 U         10       SE       1.000 0.648 0.543 ( 1.000 0.953 ( 9 U       1.000 0.953 ( 1.000 0.648 0.543 ( 9 U       1.000 0.953 ( 1.000 0.648 0.543 ( 9 U       1.000 0.953 ( 9 U         10       SE       1.000 0.953 ( 1.000 0.648 0.543 ( 9 U       1.000 0.953 ( 9 U       1.0	4	IP			1.000	0.201	0.129	0.150 0	Multiple dependent (y) variables	
6       MR       1.000 0.953 (         7       SASA       1.000 (         8       TE       1         9       DH       1         10       SE       1         11       1       1         12       1       1         13       1       1         14       1       1         15       1       1         16       1       1         17       1       1         18       1       1         19       Mariables       Normalized! \$2\$1: \$2\$30         12       1       1       1         13       1       1       1         14       1       1       1         15       1       1       1         16       1       1       1         17       1       1       1         18       Meck       New Table       Normalized Intercorrelation Sheet3         Regression Analysis       Sum=0       1         17       1       1       1         18       1       1       1         17       1       1	5	W				1.000	0.648	0.543 0		
7       SASA       1.000 (         8       TE       1         9       DH       1         10       SE       Y variables:       Normalized! \$C\$1: \$C\$30         12       1       1       1       1         13       1       1       1       1         14       1       1       1       1         15       1       1       1       1         16       1       1       1       1         17       1       1       1       1         16       1       1       1       1         17       1       1       1       1         18       1       1       1       1         19       QSAR       Regression Analysis       Sum=0	6	MR					1.000	0.953 (		
8       TE       1	7	SASA						1.000 0		
9       DH         10       SE         11       Image: Section of the sectin of the section of the section of the sectin	8	TE						1		
10     SE       11     Image: Sector of the	9	DH							X variables: Normalized \$C\$1:\$C\$30	
11   12   13   14   15   16   17   17   18   19   19   10   17   18   19   10   17   18   19   11   10   11   12   13   14   15   16   17   18   19   11   10   11   12   13   14   15   16   17   18   17   17   18   19   110   111   111   120   111   120	10	SE								
12    13    14    15    16    17    17    Mathematical Intercorrelation Sheet3      Regression Analysis      Sum=0   Imathematical Table     Y check     Normalized     Intercorrelation     Pagina 53	11			1					Y variable: Normalized!\$B\$1:\$B\$30	
13     14     15     16     16     16     16     16     17     16     17     16     17     16     17     16     17     16     17     16     17     16     17     16     17     16     17     16     17     16     17     16     17     16     17     16     17     16     16     17     16     <	12									
14     15     14       15     16     16       17     16       17     16       17     16       17     16       18     17       Initial Table Y check New Table Normalized Intercorrelation Sheet3     Regression Analysis       QSAR     Pagina 53	13									
16     17     16     QK       17     17     17     17       16     17     17     17       17     17     17     17       11     17     17     17       11     17     17       11     17     17       11     17     10       11     17       12     11       13     Regression Analysis       Sum=0     11       QSAR     Pagina 53	14									
17     Image: Concelence of the second	16									
Initial Table     Y check     New Table     Normalized     Intercorrelation     Sheet3     Regression     Analysis     Sum=0       QSAR     Pagina 53	17								Help Cancel OK	<b>_</b>
Initial Table Y check New Table Normalized Intercorrelation Sheet3 Regression Analysis Sum=0		•								
QSAR Pagina 53	Initial Tab	le Y cheo	k New	Table N	ormalized	Intercor	rrelation	Sheet3	Regression Analysis Sum=0	
QSAR Pagina 53										
QSAR Pagina 53		0.0								
		- QS	AR						Pa	igina 53

We start with the uniparametric regressions using the «Statistic → Regression» menu. Then by clicking in the X and Y variable boxes we define the areas containing the values. While doing this operation select also the label of the variables. Then click on the OK button ....

🐻 * Raje	sh.Teacher.gnum	eric - Gnumeric				
<u>File E</u> di	it <u>V</u> iew <u>I</u> nsert I	F <u>o</u> rmat <u>T</u> ools <u>S</u> tatistics	<u>D</u> ata <u>H</u> elp			
<u> </u>		🎽 🗊 🖪   🥱	• • • • •	∑ f(x) <sup>R</sup> . z. ⊻z № R	100% -	
Sans		▼ 10 ▼			1 19 18 · 42 - 5	8 +≣ +≡   ▼
A1	چ 💫	🕻 🖑 👻 = SUMM/	ARY OUTPUT			
	Α	В	С	D	E	F
1	SUMMARY OUT	TPUT	Response Variable	RS		
2						
3	Regression Sta	tistics	÷			
4	Multiple R	0.62410437320237				
5	R^2	0.38950626865032				
6	Standard Error	0.75297888457587				
7	Adjusted R^2	0.36689538971144				
8	Observations	29				
9						
10	ANOVA					
11		df	SS	MS	F	Significance of F
12	Regression	1	9.76703180885487	9.76703180885487	17.2264983463603	0.00029690051514
13	Residual	27	15.3083844166624	0.56697720061712		
14	Total	28	25.0754162255172			
15						
16		Coefficients	Standard Error	t-Statistics	p-Value	Lower 95%
17	Intercept	0.90633024200569	0.32789999371423	2.76404470686136	0.01015913570582	0.23353502854898
18	EA	2.40923773665525	0.58047183688409	4.15048170052107	0.00029690051514	1.21820790777883
19						
	<u> </u>					<u> </u>
Initial Tabl	e JY check New Ta	ble Normalized Intercorr	elation Sheet3 Regression	on (1)	Su	n=0
				1		1
	QSAR					Pagina 54
						¥

... and a new sheet «Regression (1)» will be created with all the statistical values

		🔏 🗊 🖸   <	👆 🝷 🚸 🝷 🛛 🍛	$\sum f(x) \frac{1}{Mz} \frac{z}{Ma}$	100% 🔻					
s		• 10 •			169 % · % -	8 🖅 🖅 🔛	• <u>4</u> • <u>A</u> •			
	<u>م</u>	¥ ≪8 → =								
	Α	В	С	D	E	F	G	Н	1	
7	Intercept	0.9063302420056	9 0.32789999371423	2.76404470686136	0.01015913570582	0.23353502854898	1.57912545546239			
3	EA	2.4092377366552	5 0.58047183688409	4.15048170052107	0.00029690051514	1.21820790777883	3.60026756553168			
)										
)	Constant	EA	Prediction	RS	Residual	Leverages	Internally studentized	Externally studentized	p-Value	Í.
		1 0.4316617502458	2 1.94630602016857	0.301	-1.64530602016857	0.03821898823821	-2.2281	-2.4098	2.33%	t
		1 0.4046214355948	9 1.88115947370051	0.602	-1.27915947370051	0.04120185573742	-1.7349	-1.8015	8.32%	l
		1 0.3647984267453	3 1.785216377993	0.6989	-1.086316377993	0.04717721504616	-1.4780	-1.5097	14.32%	L
		1 0.496066863323	5 2.10147324902887	0.6989	-1.40257324902887	0.0346144289285	-1.8958	-1.9928	5.69%	L
		1 0.4075712881022	6 1.88826636967884	1.301	-0.58726636967884	0.04083421996198	-0.7964	-0.7904	43.64%	Ĺ
		1 0.0717797443461	2 1.07926471081181	1.301	0.22173528918819	0.14910423874043	0.3192	0.3138	75.62%	Ĺ
		1 0.3195673549655	9 1.67624397299188	1.3979	-0.27834397299188	0.05625031936562	-0.3805	-0.3743	71.12%	t.
		1 0.519665683382	5 2.15832841685554	1.4771	-0.68122841685554	0.03452788443552	-0.9207	-0.9175	36.73%	t.
		1	0 0.90633024200569	1.602	0.69566975799431	0.18963444343223	1.0263	1.0234	31.55%	t
		1 0.3235004916420	8 1.68571983429632	1.6989	0.01318016570368	0.05536482337662	0.0180	0.0177	98.60%	ł.
		1 0.3628318584070	8 1.78047844734078	1.8808	0.10032155265922	0.0475211343019	0.1365	0.1340	89.44%	L
		1 0.5830875122910	5 2.31112668038972	2	-0.31112668038972	0.03757517487862	-0.4212	-0.4146	68.18%	Ĺ
		1 0.4326450344449	5 1.94867499549469	2.0701	0.12042501450532	0.03812689602123		0.1734	86.37%	i.
		1 0.8608652900 87	9 02 9 250 61	2.0 91	-0.5 1 5 38 01618	0.107247 3334 49	▲ -1 663	-1.2775	21.27%	Ĺ
		1 0.3805309734 1	વી/ <mark>કે</mark> વ7/ 9/ 23, 10‴ વ	2. 76	0.352 8 17 78922	0.0445 434 16	0 795	0.4724	64.06%	t.
		1 0.5471976401179	9 2.2 9 944 598 666	2.2041	-0.02055944598666	0.0352635126258	-0.0278	-0.0273	97.84%	t.
		1 0.4513274336283	2 1.9936853266908	2.301	0.3073146733092	0.03659548646297	0.4158	0.4093	68.57%	t.
		1 0.5201573254670	6 2.15951289951859	2.3424	0.18288710048141	0.03453312009996	0.2472	0.2428	81.00%	L
		1 0.3480825958702	1 1.74494396744911	2.4771	0.73215603255089	0.05024704807097	0.9977	0.9967	32.81%	L
		1 0.79006882989184	4 2.80979388153616	2.574	-0.23579388153616	0.08078153719068	-0.3266	-0.3211	75.07%	L
		1 0.3500491642084	6 1.74968189810133	2.602	0.85231810189867	0.04986865367381	1.1613	1.1676	25.36%	Ĺ
		1 0.4006882989183	9 1.87168361239606	2.8129	0.94121638760394	0.04170812517068	1.2769	1.2909	20.81%	Ĺ
		1 0.7035398230088	5 2.60132493283836	2.903	0.30167506716164	0.05652498618567	0.4125	0.4060	68.81%	t
		1 0.7723697148475	9 2.76715250566616	2.903	0.13584749433384	0.07509599196959	0.1876	0.1842	85.53%	t.
		1 0.4247787610619	5 1.92972326288579	3	1.07027673711421	0.03889581055506	1.4499	1.4792	15.11%	t
		1 0.6061946902654	9 2.36679736555334	3.3424	0.97560263444666	0.03987367119381	1.3223	1.3399	19.19%	Ľ
		1 0.9906588003933	1 3.29306280806288	3.4771	0.18403719193712	0.17123955311578	0.2685	0.2638	79.41%	L
		1	1 3.31556797866094	3.875	0.55943202133906	0.17661746218617	0.8188	0.8118	42.43%	I.
		1 0.9532940019665	7 3.20304212567065	3.875	0.67195787432935	0.15076504400564	0.9684	0.9646	34.36%	i.

QSAR

Pagina 55

By analysing the new sheet we can observe the intercept and constant values of the line and thus we can write its equation

🐻 * Ra	jesh.Teacher.gnum	eric - Gnumeric		N		
<u>F</u> ile	<u>dit V</u> iew <u>I</u> nsert i	F <u>o</u> rmat <u>T</u> ools <u>S</u> tatistics	<u>D</u> ata <u>H</u> elp	43		
		🔏 🕒 🖪   🥱	• • 🚸 •   🥹	$\sum f(\mathbf{x})  \frac{\mathbf{n}}{\mathbf{M} \mathbf{z}}  \frac{\mathbf{z}}{\mathbf{M} \mathbf{n}}$	100% -	
Sans		▼ 10 ▼			19 % · 4% _%	
A1	چ 💫	🕻 🐗 🔻 = SUMMA	ARY OUTPUT			
	Α	В	С	D	E	F
1	SUMMARY OUT	TPUT	Response Variable	RS		<b>A</b>
2						
3	Regression Sta	tistics				
4	Multiple R	0.85305842334863				
5	R^2	0.72770867364606				
6	Standard Error	0.50287390153896				
7	Adjusted R^2	0.71762380970702				
8	Observations	29				
9						
10	ANOVA					
11		df	SS	MS	F	Significance of F
12	Regression	1	18.2475978825939	18.2475978825939	72.1585018940464	4.1569053282004
13	Residual	27	6.8278183429233	0.25288216084901		
14	Total	28	25.0754162255172			
15						
16		Coefficients	Standard Error	t-Statistics	p-Value	Lower 95%
17	Intercept	0.66259068100773	0.19713001110905	3.36118624089754	0.00232964121029	0.2581133
18	IP	2.6878988550149	0.31642381929091	8.49461605336265	4.15690532820046E-009	2.0386508
19						<b>_</b>
	4					<u> </u>
<ul> <li>Reg</li> </ul>	ession (2) Regression	n (3) Regression (4) Re	gression (5) Regression	(6) •	Sum=0	
	n		~			

QSAR

Pagina 56

Reapeat for all the descriptors

🐻 * Raje	sh.Teacher.gnum	eric - Gnumeric		N		
<u>File</u> <u>E</u> o	lit <u>V</u> iew <u>I</u> nsert I	ormat <u>T</u> ools <u>S</u> tatistics	<u>D</u> ata <u>H</u> elp	2		
C E		🎽 🗊 🚺   🥱	• • 🚸 •   🥹	∑ f(x) <sup>n</sup> . z. ⊻ z № n	100% -	
Sans		▼ 10 ▼			19 % · % = (	
A1	4	SUMMA	ARY OUTPUT			
	Α	В	С	D	E	F
1	SUMMARY OUT	TPUT	Response Variable	RS		*
2						
3	Regression Sta	tistics				
4	Multiple R	0.68473965334663				
5	R^2	0.46886839286526				
6	Standard Error	0.70233331011526				
7	Adjusted R^2	0.44919685186027				
8	Observations	29				
9						
10	ANOVA					
11		df	SS	MS	F	Significance of F
12	Regression	1	11.7570701060858	11.7570701060858	23.8348583238253	4.1811541296249
13	Residual	27	13.3183461194314	0.49327207849746		
14	Total	28	25.0754162255172			
15						
16		Coefficients	Standard Error	t-Statistics	p-Value	Lower 95%
17	Intercept	1.13322035863428	0.24353781773917	4.65315969878639	7.73582994698183E-005	0.6335220
18	W	2.45877550982804	0.50363115884698	4.88209568974486	4.181154129625E-005	1.4254097
19						<b>_</b>
	4					<u> </u>
Regre	ssion (2) Regression	n (3) Regression (4) Re	egression (5) Regression	(6)	Sum=0	1.

QSAR

🐻 * Raje	sh.Teacher.gnum	eric - Gnumeric		Ν			JN
<u>File</u>	lit <u>V</u> iew <u>I</u> nsert I	Format <u>T</u> ools <u>S</u> tatistics	<u>D</u> ata <u>H</u> elp	hi			
C E		🔏 🗊 🖪   🥱	• • 🚸 •   🥹	∑ f(x) <sup>A</sup> z MA	100% -		
Sans		▼ 10 ▼			19 % · % -	20	•
A1	4	🖑 🔻 = Summa	ARY OUTPUT				
	Α	В	С	D	E	F	
1	SUMMARY OUT	TPUT	Response Variable	RS			-
2			•				
3	Regression Sta	tistics					
4	Multiple R	0.58300627675984					
5	R^2	0.33989631874137					_
6	Standard Error	0.78297567625897					
7	Adjusted R^2	0.31544803425031					
8	Observations	29					
9							
10	ANOVA						
11		df	SS	MS	F	Significance of F	
12	Regression	1	8.52304166596096	8.52304166596096	13.902665394198	0.00090322861833	
13	Residual	27	16.5523745595563	0.6130509096132			
14	Total	28	25.0754162255172				
15							
16		Coefficients	Standard Error	t-Statistics	p-Value	Lower 95%	
17	Intercept	1.01386008698312	0.33455590397855	3.03046538687935	0.00533150899806	0.32740807373112	
18	MR	2.90047981347728	0.77789469912203	3.72862781652956	0.00090322861833	1.30437173121007	
19							_
	4						•
<ul> <li>Regre</li> </ul>	ssion (2) Regression	n (3) Regression (4) Re	gression (5) Regression	(6)	Su	m=0	11.

QSAR

🐻 * Raje	sh.Teacher.gnum	eric - Gnumeric		N		
<u>F</u> ile <u>E</u> di	t <u>V</u> iew <u>I</u> nsert i	ormat <u>T</u> ools <u>S</u> tatistics	<u>D</u> ata <u>H</u> elp	43		
		🔏 🗊 🖪   🥱	• • 🚸 •   🥹	∑ f(x) <sup>n</sup> z ⊻. ⊻z № n	100% -	
Sans		▼ 10 ▼			19 % · % = %	
A1	الا 😜	SUMMA	ARY OUTPUT			
	Α	В	С	D	E	F
1	SUMMARY OUT	TPUT	Response Variable	RS		<b>A</b>
2						
3	Regression Sta	tistics				
4	Multiple R	0.5314367259195				
5	R^2	0.28242499365604				
6	Standard Error	0.81634896217295				
7	Adjusted R^2	0.25584814156922				
8	Observations	29				
9						
10	ANOVA					
11		df	SS	MS	F	Significance of F
12	Regression	1	7.08192426841418	7.08192426841418	10.6267285807022	0.0030108119412
13	Residual	27	17.9934919571031	0.66642562804085		
14	Total	28	25.0754162255172			
15						
16		Coefficients	Standard Error	t-Statistics	p-Value	Lower 95%
17	Intercept	1.37607010059851	0.27841463014938	4.94252080021867	3.55431958487801E-005	0.804810466223
18	SASA	2.39260481334891	0.73395794823225	3.2598663439936	0.00301081194124	0.8866474973523
19						
	<u> </u>					<b>)</b>
Regres	sion (2) Regression	n (3) Regression (4) Re	gression (5) Regression	(6)	Sum=0	1.

QSAR

🐻 *Ra	jesh.Teacher.gnum	eric - Gnumeric		2		
<u>F</u> ile <u>E</u>	dit <u>V</u> iew <u>I</u> nsert	F <u>o</u> rmat <u>T</u> ools <u>S</u> tatistics	<u>D</u> ata <u>H</u> elp	0		
C E		🎽 🖪 🛛 🥱	- 🚸 -   🥹 🗆	∑ f∞ <mark>n. z.</mark>   <b>111</b>	100%	
Sans		▼ 10 ▼ ▲			؉ · ‰ ≞   ∓≣	•=   •
A1	ال 😜	SUMMAR	Y OUTPUT			
	Α	В	С	D	E	F
1	SUMMARY OUT	TPUT	Response Variable	RS		<b>A</b>
2						
3	Regression Sta	tistics				
4	Multiple R	0.39548789311373				
5	R^2	0.15641067359953				
6	Standard Error	0.88513136674933				
7	Adjusted R^2	0.12516662447359				
8	Observations	29				
9						
10	ANOVA					
11		df	SS	MS	F	Significance of
12	Regression	1	3.92206274262184	3.92206274262184	5.00609485566518	0.03370951
13	Residual	27	21.1533534828954	0.78345753640353		
14	Total	28	25.0754162255172			
15						
16		Coefficients	Standard Error	t-Statistics	p-Value	Lower 95%
17	Intercept	3.70963389913321	0.72169170143323	5.14019198470223	2.08995305899929E-005	2.228844842
18	TE	-1.80394562279643	0.80625775525263	-2.23743041359171	0.0337095184943	-3.458249889
19						
	4					•
<ul> <li>Regr</li> </ul>	ession (2) Regressio	n (3) Regression (4) Regr	ression (5) Regression (6		Sum=0	1.

QSAR

🐻 *Raje	sh.Teacher.gnum	eric - Gnumeric	7			
<u>F</u> ile <u>E</u> di	t <u>V</u> iew <u>I</u> nsert I	Format <u>T</u> ools <u>S</u> tatistics	Data Help			
C 🖻		🔏 🖪 🖪   🥱	• • • •	$\sum f(\mathbf{x}) \stackrel{\mathbf{R}}{\underbrace{\mathbf{M}}} \frac{\mathbf{z}}{\mathbf{M}}$	100% -	
Sans		▼ 10 ▼			19 % · % = %	
A1	الا 😜	SUMM/	ARY OUTPUT			
	Α	В	С	D	E	F
1	SUMMARY OUT	TPUT	Response Variable	RS		<b>A</b>
2						
3	Regression Sta	tistics				
4	Multiple R	0.719230883066				
5	R^2	0.51729306315589				
6	Standard Error	0.66955146280539				
7	Adjusted R^2	0.49941502845796				
8	Observ ations	29				
9						
10	ANOVA					
11		df	SS	MS	F	Significance of F
12	Regression	1	12.9713388692068	12.9713388692068	28.9345597486613	1.1016391606563
13	Residual	27	12.1040773563104	0.44829916134483		
14	Total	28	25.0754162255172			
15						
16		Coefficients	Standard Error	t-Statistics	p-Value	Lower 95%
17	Intercept	0.63847706491587	0.30512596126205	2.09250324775717	0.0459264945109	0.0124103
18	DH	2.80306342879905	0.5211040948343	5.37908540075925	1.10163916065638E-005	1.73384614
19						<b>•</b>
						•
<ul> <li>Regres</li> </ul>	sion (3) Regression	n (4) Regression (5) Re	egression (6) Regression	(7) •	Sum=0	

QSAR

🐻 *Raje	sh.Teacher.gnum	eric - Gnumeric		N		-OX
<u>File E</u> di	it <u>V</u> iew <u>I</u> nsert I	F <u>o</u> rmat <u>T</u> ools <u>S</u> tatistics	<u>D</u> ata <u>H</u> elp	43		
		🎽 🕒 🖪   🥱	• 🚸 •   🥹 :	∑ f∞ <mark>%, z.</mark> ¥z ¥n   ∭	100%	
Sans		▼ 10 ▼ A			₩ · % - % •	-= ·
A1	الا 😜	SUMMAR	Y OUTPUT			
	Α	В	С	D	E	F
1	SUMMARY OUT	TPUT	Response Variable	RS		<b>A</b>
2						
3	Regression Sta	tistics				
4	Multiple R	0.78335040208227				
5	R^2	0.61363785244245				
6	Standard Error	0.59901745673537				
7	Adjusted R^2	0.59932814327365				
8	Observations	29				
9						
10	ANOVA					
11		df	SS	MS	F	Significance of
12	Regression	1	15.387224561727	15.387224561727	42.8826222255078	5.0549005851
13	Residual	27	9.68819166379022	0.35882191347371		
14	Total	28	25.0754162255172			
15						
16		Coefficients	Standard Error	t-Statistics	p-Value	Lower 95%
17	Intercept	3.65275893626293	0.2567617455734	14.2262583863715	4.60469169074284E-014	3.12592
18	SE	-2.99020144187781	0.45662509913122	-6.54848243683281	5.05490058511252E-007	-3.9271
19						<b>_</b>
	•					•
<ul> <li>Regres</li> </ul>	ssion (4) Regression	n (5) Regression (6) Regr	ression (7) Regression (8		Sum=0	1.

QSAR

🐻 *Raje	sh.Teach	er.gnumeric	- Gnumeric										1×
<u>F</u> ile <u>E</u> d	it <u>V</u> iew	Insert Forma	at <u>T</u> ools <u>S</u> ta	tistics <u>D</u> ata	a <u>H</u> elp								
D E		8 8 8	6 🗉 🚺	- 🧇	۰ چ	Σ 🍥	$f(x) \xrightarrow{8}_{M Z}$	2. Ma	100% 🔻				
Sans			<b>v</b> 10	- A (	AA			1 88 <b>q</b>	) % ·				-
B2		🔹 🐰 <	2	='Regression	(1)'!85								
	Α	В	С	D	E	F	G	Н	1	J	K	L	
1		EA	IP	W	MR	SASA	TE	DH	SE				-
2	RS	='Regres	sion (1)'!B5	т									
3				÷									
4													
5													
6													_
7													_
8													_
9													-
10													-
12													-
13	-												+
14													+
15	1												+
16	1												+
17	ĺ .												1
18													
19													J
Intero	orrelation	Regressions	Regression (1)	Regressio	on (2) Reg	ression (3)	•			Sum=	D		
	G QS.	AR									F	Pagina 63	

#### And record the r^2 values in a new sheet «Regressions»

🐻 * Raje	sh.Teacher.	gnumeric - (	Gnumeric							×
<u>F</u> ile <u>E</u> di	it <u>V</u> iew <u>I</u> n	sert F <u>o</u> rmat	<u>T</u> ools <u>S</u> ta	atistics <u>D</u> ata	a <u>H</u> elp					
C 🖪		A 🗳   🎉		🡆 👻	• •	i (i)	<i>f(x)</i>	ia   🛍	100% -	
Sans			<b>•</b> 10	• A (	AA			88 🦉	<b>)</b> %	•
D3		😜 🐰 🦿	· = [							
	Α	В	С	D	E	F	G	Н	1	
1		EA	IP	W	MR	SASA	TE	DH	SE	*
2	RS	0.390	0.728	0.469	0.340	0.282	0.156	0.517	0.614	
3	[									
4										
5										_
7										- 1
- / 8	I									- 1
9	l									
10										
11	i									
• Interco	orrelation	Regressions	Regression	(1)			Sum=	0		

QSAR



# Compare the Regressions with the initial correlations and the they should be the same values.

### MLR with all variables $\rightarrow$ Overfitting!

		▼ 10 ▼ A A		P	5 · 422 _23 .	• • • <u>•</u> • <u>A</u>	-	
	الا 😜	🗧 🛫 📼 🗧 SUMMARY OUTPL	т					
	Α	В	С	D	E	F	G	
1	SUMMARY OUT	TPUT	Response Variable	RS				
2								
3	Regression Sta	tistics						
+	Multiple K							
5	Standard Error	1						
,	Adjusted R^2							
3	Observations							
•								
0	ANOVA							
1		df	SS	MS	F	Significance of F		
2	Regression	9	25.0754162255172	2.7861573583908	#NUM!	#NUM!		
3	Residual	19	0	0				
.4	Total	28	25.0754162255172					
.5								
.6		Coefficients	Standard Error	t-Statistics	p-Value	Lower 95%	Upper 95%	
./	Intercept	3.63174224320954E-170	0	#DIV/0!	#DIV/0!	3.631/4224320954E-1/0	3.63174224320954E-170	
0 0	EA	-2 79620210240201E-171	0	#DIV/01	#DIV/0!	-2 78620210240201E-171	_2 78620210240201E_171	
0		-1 636235223028E-171	0	#DIV/01	#DIV/01	-1.636235223028E-171	-1.636235223028E-170	
1	W	7.30399141842223E-171	0	#DIV/0!	#DIV/0!	7.30399141842223E-171	7.30399141842223E-171	
2	MR	-5.35840666781962E-170	0	#DIV/01	#DIV/0!	-5.35840666781962E-170	-5.35840666781962E-170	
3	SASA	9.95220891031855E-171	0	#DIV/0!	#DIV/0!	9.95220891031855E-171	9.95220891031855E-171	
4	TE	-2.89869680960785E-170	0	#DIV/01	#DIV/0!	-2.89869680960785E-170	-2.89869680960785E-170	
5	DH	1.0869123178817E-170	0	#DIV/0!	#DIV/0!	1.0869123178817E-170	1.0869123178817E-170	
6	SE	-9.90380946104129E-172	0	#DIV/0!	#DIV/0!	-9.90380946104129E-172	-9.90380946104129E-172	
7								
	1 (2) Den :	- (2)   Deservation (4)   De	(5) December (6) 10	anarian (7) Dama in	(0) Demosting (0) L			
egre	ssion (2) Regressio	in (3) Regression (4) Regression	(5) Regression (6) Rej	gression (7) Regression	(8) Regression (9)		Sum=0	

Just to try, instead of highlighting one parameter, select all the 8 descriptors while doung the regressions,

in the new sheet «Regression (9)» the values are clearly indicating that 8 parameters give a perfect regression of  $r^2 = 1!$ 

#### This is overfitting!

We are using collinear parameters and the model is LYING!

#### No differences with unscaled data!

File Fi	Rajesh.Teacher.NO-Normalized.gnumeric - Gnumeric												
Sans		▼ 10 ▼			🖗 % · 🗞 🖧	· · · · ·	<u>A</u> -						
F21	ال 😜	🕻 🖑 🔻 = [{=mmu	ult(mmult(mmult('Regression	n (2)'!\$A\$21:\$B\$49,minver	se(mmult(transpose('Regression (2)'	\$A\$21:\$B\$49),'Regression (2)'!\$A\$	21:\$B\$49))),transpose('Regression (;						
	A	В	С	D	E	F	G						
1	SUMMARY OUT	TPUT	Response Variable	RS			<u> </u>						
2													
3	Regression Sta	atistics											
4	Multiple R	0.85305842334863											
5	R^2	0.72770867364606											
6	Standard Error	0.50287390153896											
7	Adjusted R^2	0.71762380970702											
8	Observations	29											
9													
10	ANOVA												
11		df	SS	MS	F	Significance of F							
12	Regression	1	18.2475978825939	18.2475978825939	72.1585018940464	4.15690532820046E-009							
13	Residual	27	6.8278183429233	0.25288216084901									
14	Total	28	25.0754162255172										
15													
16		Coefficients	Standard Error	t-Statistics	p-Value	Lower 95%	Upper 95%						
17	Intercept	-27.72040176533	3.51614143537824	-7.883756178411	1.78162061194726E-008	-34.9349280627027	-20.5058754679595						
18	IP	5.0810942438845	0.59815466784671	8.49461605336265	4.15690532820046E-009	3.85378224282119	6.30840624494782						
19													
		1		1									
NO-No	ormalized Intercorr	elation Regressions R	egression (1) Regression	n (2) 🕨		Sum=0.	14879254218158						
		^											
	OSAR												

IF we repeated with unscaled data (without normalizing) we can observe that the results are the same, the only differences are the intercept and the variable coefficient values. Although it seems that scaling is not necessary, apply always a normalization in you matrix!



Going back to the intercorrelation data it is possible to see that the first parameter to use is «IP» and that we can associate with it only «W», «MR», «SASA», «TE» and «DH» values.

The ones that display low correlation with «IP».

At first glance we could expect a godd correlation by using atriparametric model including «IP», «W» and «DH». But, this can be confirmed by a stepwise regression.

М	N	0 1	P	Q	R	S	Т	U	V	W	Х	Y	Z	
IP	w	I	P	MR		IP	SASA		IP	TE		IP	DH	
5.59	1.88	5.	59	71.83		5.59	135.70		5.59	-1682.34		5.59	-440.50	
5.62	1.83	5.	62	77.93		5.62	152.92		5.62	-2451.02		5.62	-364.00	
5.60	1.75	5.	60	71.83		5.60	137.37		5.60	-1682.34		5.60	-441.14	
5.59	2.01	5.	59	74.88		5.59	140.25		5.59	-2066.68		5.59	-400.95	
5.72	1.86	5.	72	71.83		5.72	135.59		5.72	-1682.34		5.72	-438.21	
5.61	1.28	5.	61	71.83		5.61	133.58		5.61	-1682.34		5.61	-438.67	
5.87	1.73	5.	87	77.93		5.87	152.60		5.87	-2451.01		5.87	-359.04	
5.89	2.13	5.	89	74.88		5.89	142.07		5.89	-2066.68		5.89	-401.52	
5.98	1.27	5.	98	63.70		5.98	131.21		5.98	-1370.07		5.98	-479.38	
5.73	1.71	5.	73	74.88		5.73	139.27		5.73	-2066.69		5.73	-402.63	
5.89	1.82	5.	89	74.88		5.89	139.82		5.89	-2066.69		5.89	-399.69	
5.89	2.27	5.	89	77.93		5.89	146.29		5.89	-2451.03		5.89	-362.81	
5.79	1.93	5.	79	74.88		5.79	140.65		5.79	-2066.68		5.79	-402.22	
5.85	3.00	5.	85 1	.02.92		5.85	183.21		5.85	-21831.26		5.85	-269.02	
6.02	1.88	6.	02	76.83		6.02	145.35		6.02	-2375.81		6.02	-323.53	
5.83	2.18	5.	83	77.93		5.83	145.48		5.83	-2451.03		5.83	-365.01	
5.99	2.01	5.	99	80.98		5.99	157.94		5.99	-2835.36		5.99	-324.99	
5.91	2.13	5.	91	77.93		5.91	147.35		5.91	-2451.03		5.91	-365.53	
5.92	1.80	5.	92	82.68		5.92	152.62		5.92	-2490.34		5.92	-358.42	
6.01	2.82	6.	01	80.91		6.01	148.48		6.01	-4564.72		6.01	-353.60	
5.92	1.80	5.	92	76.38		5.92	146.25		5.92	-2375.80		5.92	-323.36	
6.02	1.92	6.	02	77.93		6.02	145.05		6.02	-2451.03		6.02	-363.68	
5.97	2.58	5.	97	83.89		5.97	149.74		5.97	-6678.40		5.97	-341.98	
5.92	2.75	5.	92	86.88		5.92	156.43		5.92	-8792.08		5.92	-331.89	
6.12	1.99	6.	12	76.23		6.12	143.95		6.12	-2475.06		6.12	-364.74	
6.11	2.37	6.	11	80.98		6.11	151.35		6.11	-2835.38		6.11	-326.35	
6.03	3.44	6.	03	83.96		6.03	155.36		6.03	-4949.06		6.03	-315.98	
6.05	3.47	6.	05	89.31		6.05	162.62		6.05	-9295.45		6.05	-296.47	
5.97	3.31	5.	97	92.91		5.97	166.51		5.97	-11290.11		5.97	-284.65	

QSAR

Pagina 69

#### Let's prepare the biparametric combinations

File E	Rajesh.Teacher.NO-Normalized.gnumeric - Gnumeric       Ele Edit View Insert Format Iools Statistics Data Help											
			) - 🛞 - 🥹	∑ <i>f(x)</i> <mark>⊻ z № </mark>	100% -							
Sans		▼ 10 ▼	AAA		19 % · 🗞 🖧	📮 🖅 🗄 👻 📥 🗸	<u>A</u> -					
F58		《 ≪ → =										
	А	В	С	D	E	F	G					
1	SUMMARY OU	TPUT	Response Variable	RS			A					
2												
3	Regression Sta	atistics										
4	Multiple R	0.91769207990091										
5	R^2	0.84215875351285										
6	Standard Error	0.39016441077004										
/	Adjusted R^2	0.83001711916768										
8	Observations	29										
10												
10	ANOVA	df	cc	MC	E	Significance of F						
12	Regression	2	21 1174912722075	10 5597406361497	60 361 2350 340 770	3 77456166623094E-011						
13	Residual	2	3 05703405321075	0 15222826743153	09.3012330340779	3.77430100033094E-011						
14	Total	20	25 0754162255172	0.13222020743133								
15	/ ocur	20	2010704102200172									
16		Coefficients	Standard Error	t-Statistics	p-Value	Lower 95%	Upper 95%					
17	Intercept	-23.12151769879	2.92646506772578	-7.900835022356	2.23320906535491E-008	-29.1369527966614	-17.10608260092					
18	IP	4.07088235004434	0.51914507766559	7.84151198803542	2.56892074143486E-008	3.00376435997592	5.13800034011275					
19	W	0.61656170789146	0.14200122009646	4.34194655139328	0.00019078662105	0.32467401965991	0.908449396123					
20							-					
	4	•			• •							
<ul> <li>Regn</li> </ul>	ession (4) Regressi	on (5) Regression (6)	Regression (7) Regress	ion (8) Regression (9)	Sheet1 Sheet2 Regression	(10) 🕨 Sun	n=0 //					
							10					
							D : 70					
		JSAR					Pagina 70					

And by using the «Statistic  $\rightarrow$  Regression» menu the five biparametric are easily build.

	t <u>v</u> iew <u>i</u> nsert	Pormat <u>T</u> ools <u>S</u> tatistics	<u>D</u> ata <u>H</u> eip				
) 🗎		🔏 🗊 🚺   🤶	) 🔻 🚸 👻 🛛 🥹	∑ f(x) <sup>A</sup> z Ma	100% -		
IS		▼ 10 ▼			69 % · % …	🗐 🖅 🔛 🛨 💁 🛨	<u>A</u> -
	ا 😜	🕻 🖑 - = SUMM	ARY OUTPUT				
	А	В	С	D	E	F	G
1	SUMMARY OUT	TPUT	Response Variable	RS			
2							
3	Regression Sta	tistics					
4	Multiple R	0.90317927777274					
5	R^2	0.81573280779809					
6	Standard Error	0.42156195553287					
7	Adjusted R^2	0.80155840839794					
8	Observations	29					
9							
10	ANOVA						
11		df	SS	MS	F	Significance of F	
12	Regression	2	20.454839684347	10.2274198421735	57.5497264307116	2.82373695431237E-010	
13	Residual	26	4.62057654117024	0.1777144823527			
14	Total	28	25.0754162255172				
15							
16		Coefficients	Standard Error	t-Statistics	p-Value	Lower 95%	Upper 95%
1/	Intercept	-26.90950309644	2.956567694227	-9.101602222397	1.44960940692087E-009	-32.9868150292698	-20.83219116362
18	IP	4.40181436773323	0.53/2051748974	8.19391653956787	1.12563128581626E-008	3.29/57331614008	5.50605541932638
.9	мк	0.04031839293135	0.01144035937905	3.5242243355/411	0.00159489291772	0.01680239743906	0.06383438842364
:0							
	1	10 la		the late of the			

QSAR

File Edi	Rajesh.Teacher.NO-Normalized.gnumeric - Gnumeric												
Sans 🔻 10 🔻 🗚 🔺 🖹 🗏 🖽 🖽 🖽 🖽 🐯 ‰ - 比 🖧 📮 🖅 🔛 🗸 🖌													
A1 😜 🔏 🥪 👻 = JSUMMARY OUTPUT													
	A	B	С	D	E	F	G						
1	SUMMARY OUT	TPUT	Response Variable	RS				-					
2													
3	Regression Sta	tistics						_					
4	Multiple R	0.8804930569455											
5	R^2	0.77526802332922						_					
6	Standard Error	0.46555375853357						_					
/	Adjusted R^2	0.7579809482007											
8	Observations	29											
9								-					
10	ANOVA	df	CC	MC	r.	Cignificance of C		-					
12	Regrossion	2	33	MS 0 72000410565714	r 44.9466956056025	2 7204 274014 22885-000							
13	Regidual	2	5 62524705420205	9.72008418303714	44.8400830030023	5.72945749145568E-009		-					
14	Total	20	25 075/162255172	0.21074030200473				-					
15	/ otal	20	20.0704102200172										
16	1	Coefficients	Standard Error	t-Statistics	p-Value	Lower 95%	Upper 95%						
17	Intercept	-27.54014761432	3.25610239717114	-8.458010300367	6.13079267350096E-009	-34.2331619469472	-20.84713328170						
18	IP	4.53544760681066	0.60063634325228	7.55107088967087	5.13335982573061E-008	3.30082192133651	5.77007329228482						
19	SASA	0.02047999002363	0.00873087054483	2.34569850949738	0.02689471437376	0.00253342859374	0.03842655145351						
20								-					
	1		1		• •			▶					
<ul> <li>Regres</li> </ul>	sion (6) Regressio	on (7) Regression (8) F	Regression (9) Sheet1	Sheet2 Regression (10	) Regression (11) Regression	(12) • Sum	n=0						

QSAR
🐻 Rajes	h.Teacher.NO-Nori	malized.gnumeric - Gnum	neric				<u> </u>
<u>File E</u> di	it <u>V</u> iew <u>I</u> nsert F	Format <u>T</u> ools <u>S</u> tatistics	Data Help				
C 🖻		🔏 🗊 🖪   🥱	- 🚸 -   🕥 :	∑ f(x) <sup>A</sup> z Ma <b>b</b> 1	100% -		
Sans		• 10 • A			₩ · % - % F	+=   E • <u>4</u> • <u>A</u>	•
A1	چ 💫	SUMMAR	Y OUTPUT				
	A	В	С	D	E	F	G
1	SUMMARY OUT	TPUT	Response Variable	RS			*
2							
3	Regression Sta	tistics					
4	Multiple R	0.88170499144918					
5	R^2	0.77740369194639					
6	Standard Error	0.46333635708462					
7	Adjusted R^2	0.76028089901919					
8	Observations	29					
9							
10	ANOVA						
11		df	SS	MS	F	Significance of F	
12	Regression	2	19.4937211508096	9.7468605754048	45.4016874029613	3.29407467335866E-009	
13	Residual	26	5.58169507470764	0.21468057979645			
14	Total	28	25.0754162255172				
15							
16		Coefficients	Standard Error	t-Statistics	p-Value	Lower 95%	Upper 959
17	Intercept	-26.2642533135557	3.29558699475141	-7.96952207767067	1.9000194819616E-008	-33.0384293988774	-19.49007722823
18	IP	4.79872863573448	0.56344968096452	8.51669421042164	5.3632217820068E-009	3.6405412293177	5.9569160421512
19	TE	-5.08066165E-005	2.10880368E-005	-2.40926251133444	0.02336381837501	-9.415369698587E-005	-7.45953606E-00
20							<b>•</b>
<ul> <li>Regres</li> </ul>	sion (7) Regressio	n (8) Regression (9) She	eet1 Sheet2 Regressio	on (10) Regression (11)	Regression (12) Regression (13	) • Sum=0	

QSAR

<u>File E</u> o	lit <u>V</u> iew <u>I</u> nsert	Format <u>T</u> ools <u>S</u> tatistics	: <u>D</u> ata <u>H</u> elp				
	8 🖬   占 🕰	🔏 🖪 🗍 🔄	• • 🚸 •   🕥	∑ f(x) <sup>A</sup> Z MA	100% 🔻		
Sans		▼ 10 ▼			167 % · 4% _%	. · · · <u>*</u> ·	<u>A</u> -
A1	ډ 😜	🕻 🐗 👻 = Summ	ARY OUTPUT				
	Α	В	С	D	E	F	G
1	SUMMARY OU	TPUT	Response Variable	RS			
2							
3	Regression Sta	atistics					
4	Multiple R	0.89582250409996					
5	R^2	0.80249795885192					
6	Standard Error	0.43643864169934					
7	Adjusted R^2	0.78730549414822					
8	Observations	29					
9							
10	ANOVA						
11		df	SS	MS	F	Significance of F	
12	Regression	2	20.1229703383399	10.0614851691699	52.8221045434813	6.95713848758163E-010	
13	Residual	26	4.95244588717739	0.19047868796836			
14	Total	28	25.0754162255172				
15							
16		Coefficients	Standard Error	t-Statistics	p-Value	Lower 95%	Upper 95%
17	Intercept	-18.57278677660	4.22036925011418	-4.400749241574	0.00016342050294	-27.2478800121582	-9.897693541049
18	IP	3.9136043985866	0.63870094037851	6.12744424059761	1.77378147024894E-006	2.60073581314938	5.22647298402382
19	DH	0.00623383839961	0.00198671224603	3.13776613199662	0.00420172888299	0.00215009289178	0.01031758390743
20	ļ.,						
	4						
Regre	ssion (8) Regressio	on (9) Sheet1 Sheet2	Regression (10) Regres	ssion (11) Regression (1	2) Regression (13) Regression	(14) • Sun	1=0

QSAR

#### **Biparametric models**



The best combination resulted to be IP + W, but also IP + MR should be analyzed with other descriptor

3	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR
	IP	w	MR		IP	w	SASA		IP	w	TE		IP	w	DH	
	5.59	1.88	71.83		5.59	1.88	135.70		5.59	1.88	-1682.34		5.59	1.88	-440.50	
	5.62	1.83	77.93		5.62	1.83	152.92		5.62	1.83	-2451.02		5.62	1.83	-364.00	
	5.60	1.75	71.83		5.60	1.75	137.37		5.60	1.75	-1682.34		5.60	1.75	-441.14	
	5.59	2.01	74.88		5.59	2.01	140.25		5.59	2.01	-2066.68		5.59	2.01	-400.95	
	5.72	1.86	71.83		5.72	1.86	135.59		5.72	1.86	-1682.34		5.72	1.86	-438.21	
	5.61	1.28	71.83		5.61	1.28	133.58		5.61	1.28	-1682.34		5.61	1.28	-438.67	
	5.87	1.73	77.93		5.87	1.73	152.60		5.87	1.73	-2451.01		5.87	1.73	-359.04	
	5.89	2.13	74.88		5.89	2.13	142.07		5.89	2.13	-2066.68		5.89	2.13	-401.52	
	5.98	1.27	63.70		5.98	1.27	131.21		5.98	1.27	-1370.07		5.98	1.27	-479.38	
	5.73	1.71	74.88		5.73	1.71	139.27		5.73	1.71	-2066.69		5.73	1.71	-402.63	
	5.89	1.82	74.88		5.89	1.82	139.82		5.89	1.82	-2066.69		5.89	1.82	-399.69	
	5.89	2.27	77.93		5.89	2.27	146.29		5.89	2.27	-2451.03		5.89	2.27	-362.81	
	5.79	1.93	74.88		5.79	1.93	140.65		5.79	1.93	-2066.68		5.79	1.93	-402.22	
	5.85	3.00	102.92		5.85	3.00	183.21		5.85	3.00	-21831.26		5.85	3.00	-269.02	
	6.02	1.88	76.83		6.02	1.88	145.35		6.02	1.88	-2375.81		6.02	1.88	-323.53	
	5.83	2.18	77.93		5.83	2.18	145.48		5.83	2.18	-2451.03		5.83	2.18	-365.01	
	5.99	2.01	80.98		5.99	2.01	157.94		5.99	2.01	-2835.36		5.99	2.01	-324.99	
	5.91	2.13	77.93		5.91	2.13	147.35		5.91	2.13	-2451.03		5.91	2.13	-365.53	
	5.92	1.80	82.68		5.92	1.80	152.62		5.92	1.80	-2490.34		5.92	1.80	-358.42	
	6.01	2.82	80.91		6.01	2.82	148.48		6.01	2.82	-4564.72		6.01	2.82	-353.60	
	5.92	1.80	76.38		5.92	1.80	146.25		5.92	1.80	-2375.80		5.92	1.80	-323.36	
	6.02	1.92	77.93		6.02	1.92	145.05		6.02	1.92	-2451.03		6.02	1.92	-363.68	
	5.97	2.58	83.89		5.97	2.58	149.74		5.97	2.58	-6678.40		5.97	2.58	-341.98	
	5.92	2.75	86.88		5.92	2.75	156.43		5.92	2.75	-8792.08		5.92	2.75	-331.89	
	6.12	1.99	76.23		6.12	1.99	143.95		6.12	1.99	-2475.06		6.12	1.99	-364.74	
	6.11	2.37	80.98		6.11	2.37	151.35		6.11	2.37	-2835.38		6.11	2.37	-326.35	
	6.03	3.44	83.96		6.03	3.44	155.36		6.03	3.44	-4949.06		6.03	3.44	-315.98	
	6.05	3.47	89.31		6.05	3.47	162.62		6.05	3.47	-9295.45		6.05	3.47	-296.47	
	5.97	3.31	92.91		5.97	3.31	166.51		5.97	3.31	-11290.11		5.97	3.31	-284.65	

QSAR

Pagina 76

#### So prepare the triparametric combination for IP + W

IP	MR	SASA	IP	MR	TE	I	Р	MR	DH	
5.59	71.83	135.70	5.59	71.83	-1682.34	5	5.59	71.83	-440.50	
5.62	77.93	152.92	5.62	77.93	-2451.02	5	5.62	77.93	-364.00	
5.60	71.83	137.37	5.60	71.83	-1682.34	5	5.60	71.83	-441.14	
5.59	74.88	140.25	5.59	74.88	-2066.68	5	5.59	74.88	-400.95	
5.72	71.83	135.59	5.72	71.83	-1682.34	5	5.72	71.83	-438.21	
5.61	71.83	133.58	5.61	71.83	-1682.34	5	5.61	71.83	-438.67	
5.87	77.93	152.60	5.87	77.93	-2451.01	5	5.87	77.93	-359.04	
5.89	74.88	142.07	5.89	74.88	-2066.68	5	5.89	74.88	-401.52	
5.98	63.70	131.21	5.98	63.70	-1370.07	5	5.98	63.70	-479.38	
5.73	74.88	139.27	5.73	74.88	-2066.69	5	5.73	74.88	-402.63	
5.89	74.88	139.82	5.89	74.88	-2066.69	5	5.89	74.88	-399.69	
5.89	77.93	146.29	5.89	77.93	-2451.03	5	5.89	77.93	-362.81	
5.79	74.88	140.65	5.79	74.88	-2066.68	5	5.79	74.88	-402.22	
5.85	102.92	183.21	5.85	102.92	-21831.26	5	5.85	102.92	-269.02	
6.02	76.83	145.35	6.02	76.83	-2375.81	e	6.02	76.83	-323.53	
5.83	77.93	145.48	5.83	77.93	-2451.03	5	5.83	77.93	-365.01	
5.99	80.98	157.94	5.99	80.98	-2835.36	5	5.99	80.98	-324.99	
5.91	77.93	147.35	5.91	77.93	-2451.03	5	5.91	77.93	-365.53	
5.92	82.68	152.62	5.92	82.68	-2490.34	5	5.92	82.68	-358.42	
6.01	80.91	148.48	6.01	80.91	-4564.72	e	6. <b>01</b>	80.91	-353.60	
5.92	76.38	146.25	5.92	76.38	-2375.80	5	5.92	76.38	-323.36	
6.02	77.93	145.05	6.02	77.93	-2451.03	e	6.02	77.93	-363.68	
5.97	83.89	149.74	5.97	83.89	-6678.40	5	5.97	83.89	-341.98	
5.92	86.88	156.43	5.92	86.88	-8792.08	5	5.92	86.88	-331.89	
6.12	76.23	143.95	6.12	76.23	-2475.06	e	5.12	76.23	-364.74	
6.11	80.98	151.35	6.11	80.98	-2835.38	e	5.11	80.98	-326.35	
6.03	83.96	155.36	6.03	83.96	-4949.06	e	6.03	83.96	-315.98	
6.05	89.31	162.62	6.05	89.31	-9295.45	6	6.05	89.31	-296.47	
5.97	92.91	166.51	5.97	92.91	-11290.11	5	5.97	92.91	-284.65	

QSAR

Pagina 77

#### And the other threes for IP + MR

🐻 Rajes	h.Teacher.NO-Nori	malized.gnumeric - Gnu	ımeric					
<u>File</u> Edi	t <u>V</u> iew <u>I</u> nsert F	Format <u>T</u> ools <u>S</u> tatistics	<u>D</u> ata <u>H</u> elp					
		🔏 🗊 🗊   🦂	• 🚸 •   🕹	$\sum_{x} f(x) = \begin{cases} n, & z, \\ \underline{M} z, & \underline{M} n \end{cases}$	100% -			
Sans		▼ 10 ▼			19 % · 4% _%	= =   = • <u>4</u> •	<u>A</u> •	
A1	الا 😜	🕻 🖑 👻 = SUMM/	ARY OUTPUT					
	A	В	С	D	E	F	G	
1	SUMMARY OUT	ΓΡυτ	Response Variable	RS				<b></b>
2								
3	Regression Sta	itistics						
4	Multiple R	0.91937513026079						
5	R^2	0.84525063014204						
6	Standard Error	0.39397486517924						
7	Adjusted R^2	0.82668070575908						
8	Observations	29						
9								
10	ANOVA							
11		df	SS	MS	F	Significance of F		
12	Regression	3	21.1950113656923	7.06500378856409	45.5171821226071	2.82079971524628E-010		
13	Residual	25	3.88040485982496	0.155216194393				
14	Total	28	25.0754162255172					
15								
16		Coefficients	Standard Error	t-Statistics	p-Value	Lower 95%	Upper 95%	
17	Intercept	-23.80208339730	3.10798434250768	-7.658366572753	5.15812132000212E-008	-30.2030969720504	-17.4010698225512	
18	IP	4.07242818164555	0.52421975621695	7.76855151555969	3.99603520849363E-008	2.99277738360157	5.15207897968954	
19	W	0.49322499897487	0.22586418121859	2.18372384817194	0.03857769116353	0.02804901006902	0.95840098788072	
20	MR	0.01190276432794	0.01684151656373	0.70675133577771	0.48625672532005	-0.0227829883219	0.04658851697778	
21								
	<u>∢</u>							<u> </u>
Regres	sion (8) Regression	1 (9) Sheet1 Sheet2 Re	gression (10) Regression (	(11) Regression (12) Re	gression (13) Regression (14) Regression (14)	egression (15)	Sum=0	1.
		QSAR					Pagina 78	

#### And build the seven triparametric models

🐻 Rajes	h.Teacher.NO-Nor	malized.gnumeric - Gnum	eric				
<u>Eile E</u> di	t <u>V</u> iew <u>I</u> nsert I	Format <u>T</u> ools <u>S</u> tatistics	Data Help				
		🔏 🗊 🖪   🥱	• 🚸 •   🥹 :	$\sum f(\mathbf{x}) \stackrel{\mathbf{n}}{\overset{\mathbf{n}}{\overset{\mathbf{n}}}} \frac{\mathbf{z}}{\mathbf{M}\mathbf{n}} \qquad $	100% -		
Sans		▼ 10 ▼ A			9 % · ** =*   =	······································	•
A1	الا 😜	SUMMAR	Y OUTPUT				
	Α	В	С	D	E	F	G
1	SUMMARY OUT	TPUT	Response Variable	RS			A
2							
3	Regression Sta	tistics					
4	Multiple R	0.91787211162008					
5	R^2	0.84248921328991					
6	Standard Error	0.39747445385907					
7	Adjusted R^2	0.8235879188847					
8	Observations	29					
9							
10	ANOVA						
11		df	SS	MS	F	Significance of F	
12	Regression	3	21.125767688753	7.04192256291767	44.5730961715316	3.51350437540209E-010	
13	Residual	25	3.94964853676424	0.15798594147057			
14	Total	28	25.0754162255172				
15							
16		Coefficients	Standard Error	t-Statistics	p-Value	Lower 95%	Upper 95%
17	Intercept	-22.9105051394135	3.12042325720307	-7.34211459504019	1.08270184063941E-007	-29.3371371385327	-16.4838731402944
18	IP	4.0823438909047	0.53123426734906	7.68463960594302	4.85281976176919E-008	2.98824643675543	5.17644134505398
19	W	0.64761253157309	0.19826556494765	3.26638935885852	0.00315651362068	0.23927695687989	1.05594810626628
20	SASA	-0.00233972457511	0.01021622807493	-0.22902039362741	0.82071627039116	-0.02338044015915	0.01870099100894
21							
			1				<b>_</b>
<ul> <li>Regres</li> </ul>	sion (14) Regres	sion (15) Regression (16)	Regression (17) Reg	pression (18) Regression (	19) Regression (20) Regressi	on (21)	Sum=0
							,

QSAR

🐻 Rajes	h.Teacher.NO-Nor	malized.gnumeric - Gnu	imeric					- 🗆 🗵
<u>File E</u> di	it <u>V</u> iew <u>I</u> nsert I	Format <u>T</u> ools <u>S</u> tatistics	<u>D</u> ata <u>H</u> elp					
		🔏 🗊 🗊   🦂	• • • •	∑ f(x) <sup>n</sup> z. Mz Ma	100% -			
Sans		▼ 10 ▼			10 % · 4% _%	<u>4</u>	<u>A</u> -	
A1	4	🕻 🖑 👻 = SUMM	ARY OUTPUT					
	Α	В	С	D	E	F	G	
1	SUMMARY OUT	TPUT	Response Variable	RS				*
2								
3	Regression Sta	atistics						
4	Multiple R	0.91779372625134						
5	R^2	0.84234532394632						
6	Standard Error	0.3976559629602						
7	Adjusted R^2	0.82342676281988						
8	Observations	29						
9								
10	ANOVA	10			_			
11		df	55	MS	F	Significance of F		
12	Regression	3	21.1221596035722	7.04071986785741	44.5248091710865	3.55355920184728E-010		
13	Residual	25	3.95325662194501	0.1581302648778				
14	liotai	28	25.0754162255172					
15	1	Coefficients	Ctandard Frees	t Ctatistics	n Maha	Lower 05%	Upper 05%	
10	Intercent	22.0602524.0550	2 00062054270441	1-Statistics	4 750055175220415 000	20.2441444060226	16 0025622142570	
10	Intercept	4 05602702107155	2.99602634579441	7 56572412155011	4.75995517552941E=008	29.2441444909220	5 16016942177042	
10		4.03003702187133	0.33010020013174	2 20006240901402	0.00262421772472	0.22021000607042	1 05102520710271	
20		4 20006886E-006	2/0/60/5/5=005	0 17200277900545	0.00303431772472	-4 709922600959E-005	5.567016462E=005	
20	12	4.29090880E-000	2.49409434E-003	0.17200377809343	0.80481930103994	-4.708822090838E-003	3.307010402E-003	-
21	4				· · · · · ·			• • Ē
• Regres	sion (14) Regres	sion (15) Regression (1	6) Regression (17) R	egression (18) Regress	ion (19) Regression (20) Reg	ression (21)	Sum=0	

QSAR

🐻 Rajes	h.Teacher.NO-Nor	malized.gnumeric - Gnu	meric					_ 🗆 🗵
<u>File E</u> di	it <u>V</u> iew <u>I</u> nsert	F <u>o</u> rmat <u>T</u> ools <u>S</u> tatistics	<u>D</u> ata <u>H</u> elp					
D B		🔏 🖪 🖪   🦂	• • • • •	$\sum f(\mathbf{x}) \begin{array}{c} \mathbf{a} & \mathbf{z} \\ \mathbf{M} \mathbf{z} & \mathbf{M} \mathbf{a} \end{array}$	100% -			
Sans		▼ 10 ▼			10 % · % _ %	• • • • • •	<u>A</u> •	
A1	في 😜	🕻 🖑 – = SUMM/	ARY OUTPUT					
	Α	B	С	D	E	F	G	
1	SUMMARY OU	TPUT	Response Variable	RS				*
2								
3	Regression Sta	atistics						
4	Multiple R	0.91997255702248						
5	R^2	0.84634950567447						
0	Standard Error	0.3925/356500249						
	Adjusted R^2	0.82/91144635541						
8	Observations	29						
10								
10	ANOVA	df	CC	MC	F	Significance of F		
12	Pagrossian	2	21 2225661270402	7 07410070001607	45 0022005125542	2 50102440071272E_010		
13	Regidual	25	2 05205000046005	0 15411400202076	45.9023095123545	2.36192446671272E-010		
14	Total	23	25 0754162255172	0.13411400333070				
15	10101	20	25.0754102255172					
16		Coefficients	Standard Error	t-Statistics	n-Value	Lower 95%	Upper 95%	
17	Intercept	-21.00497066704	3.90386643144781	-5.380555671125	1.39678266422564E-005	-29.0451340874092	-12,9648072466733	
18	IP	3.87297837817665	0.57470840268017	6.73903210761297	4.61167922794022E-007	2.68934426626531	5.05661249008799	
19	W	0.51135155559572	0.19143621517046	2.67113281121017	0.01310353726964	0.11708129005888	0.90562182113255	
20	DH	0.00197715383419	0.00239437102872	0.82575081742642	0.41675629700302	-0.00295414560906	0.00690845327743	
21								-
								<u> </u>
<ul> <li>Regres</li> </ul>	sion (14) Regres	sion (15) Regression (16	5) Regression (17) R	egression (18) Regress	ion (19) Regression (20) Reg	ression (21)	Sum=0	1.

QSAR

🐻 Rajes	h.Teacher.NO-Nor	malized.gnumeric - Gnum	ieric				_	
<u>F</u> ile <u>E</u> di	it <u>V</u> iew <u>I</u> nsert i	F <u>o</u> rmat <u>T</u> ools <u>S</u> tatistics	<u>D</u> ata <u>H</u> elp					
		🔏 🗊 🗊   🥱	- 🚸 - 🛛 🕹	∑ f(x) <sup>n</sup> z.   ∭.	100%			
Sans		▼ 10 ▼ #			9 % · • * = =	🗉   🖽 🝷 💁 👻 🗛	•	
A1	4	SUMMAR	Y OUTPUT					
	Α	В	С	D	E	F	G	
1	SUMMARY OUT	TPUT	Response Variable	RS				*
2								
3	Regression Sta	tistics						
4	Multiple R	0.93029450784265						
5	R^2	0.86544787132221						
6	Standard Error	0.36736633654297						
7	Adjusted R^2	0.84930161588087						
8	Observations	29						
9								
10	ANOVA	1 -						
11		df	SS	MS	F	Significance of F		
12	Regression	3	21.7014655948922	7.23382186496407	53.6005313719128	4.96403320051664E-011		
13	Residual	25	3.37395063062503	0.134958025225				
14	Total	28	25.0754162255172					
15								
16		Coefficients	Standard Error	t-Statistics	p-Value	Lower 95%	Upper 95%	
17	Intercept	-25.5541806822902	2.61478082168975	-9.77297235405616	5.08641309667451E-010	-30.9394225915613	-20.168938773019	
18	IP	4.63976084281589	0.47464398225434	9.77524421731669	5.06281839926264E-010	3.66221326253052	5.61730842310127	
19	MR	0.13889262202855	0.03393123830337	4.09335553234888	0.00038969353447	0.06900992860008	0.20877531545701	
20	SASA	-0.07126526122011	0.02344818134093	-3.03926603875767	0.00549348056003	-0.11955769468371	-0.02297282775651	
21								
	11 	a tat la contrat	la i tat la			5 A 1 .		<u> </u>
<ul> <li>Regres</li> </ul>	sion (14) Regres	sion (15) Regression (16)	Regression (17) Reg	gression (18) Regression (	19) Regression (20) Regressi	ion (21)	Sum=0	//

QSAR

🐻 Rajes	h.Teacher.NO-Nor	malized.gnumeric - Gnumeric					
Eile Edi	it <u>V</u> iew <u>I</u> nsert	Format Tools Statistics Data	Help				
<u> </u>		🎽 🗊 🗊   🥱 🝷	🗼 🔻 🛛 🎱 🗵 f(x) 👫	2. Ma 100% -			
Sans		<ul><li>▼ 10</li><li>▼ ▲ ▲</li></ul>		- 88   ® % ·	\$	• <u>h</u> • <u>A</u> •	
A1	في 💫	🛯 🚽 🚽 = SUMMARY OUT	PUT				
	Α	В	С	D	E	F	G
1	SUMMARY OUT	TPUT	Response Variable	RS			*
2							
3	Regression Sta	tistics					
4	Multiple R	0.90928297517243					
6	Standard Error	0.82079332893842	·				
7	Adjusted B^2	0.90601099241103					
8	Observations	29					
9							
10	ANOVA						
11		df	SS	MS	F	Significance of F	
12	Regression	3	20.7322420215277	6.91074734050922	39.7793584595404	1.14188494827146E-009	
13	Residual	25	4.34317420398958	0.17372696815958			
14	Total	28	25.0754162255172				
15							
16		Coefficients	Standard Error	t-Statistics	p-Value	Lower 95%	Ur
17	Intercept	-27.9060560150851	3.0277237249918	-9.21684359267643	1.61928437395792E-009	-34.1417697537929	-21.670342
18	IP	4.20517239877165	0.55347131795757	7.5978144889055	5.93895741308059E-008	3.06527688159461	5.3450679
19	MR	0.07043044183343	0.02637803064221	2.67004170207962	0.01313653540738	0.01610387078008	0.1247570
20	TE	5.59017392867349E-005	4.42388554946174E-005	1.26363439247511	0.21801810719595	-3.52098891341316E-005	0.0001470
21	1						<b>_</b>
Regress	sion (14) Regres	sion (15) Regression (16) Re	gression (17) Regression (18)	Regression (19) Regress	ion (20) Regression (21)	Sum=0	
Regres	Sourt (1-1) [Regres	sion (15) [Regression (16) [Re	gression (17) jregression (16)	(Regression (19) Regress	ion (20) [Regression (21)]	Sum=0	

QSAR

🐻 Rajes	h.Teacher.NO-Nori	malized.gnumeric - Gnun	neric				_	
<u>F</u> ile <u>E</u> di	t <u>V</u> iew <u>I</u> nsert F	Format <u>T</u> ools <u>S</u> tatistics	<u>D</u> ata <u>H</u> elp					
		🔏 🗊 🖪   🥱	- 🚸 -   🥹 🗉	∑ f(x) <sup>n</sup> , z, ⊻z ⊻n   111	100% -			
Sans		▼ 10 ▼ ▲			19 % · •% _*   .=	+=   = • <u>\$</u> • <u>A</u>	•	
A1	۵ 🖇	SUMMAR						
	A	В	С	D	E	F	G	
1	SUMMARY OUT	TPUT	Response Variable	RS				*
2								
3	Regression Sta	tistics						
4	Multiple R	0.90353582275065	1					
5	R^2	0.8163769829937						
6	Standard Error	0.42915841271113	i					
7	Adjusted R^2	0.79434222095294						
8	Observations	29	-					
9	<u> </u>							
10	ANOVA							
11	<u> </u>	df	SS	MS	F	Significance of F		
12	Regression	3	20.4709926454989	6.82366421516631	37.0495030299709	2.35619727643141E-009		
13	Residual	25	4.60442358001831	0.18417694320073				
14	Total	28	25.0754162255172					
15	<u> </u> '							
16	<u> </u>	Coefficients	Standard Error	t-Statistics	p-Value	Lower 95%	Upper 95%	
17	Intercept	-25.2420591257467	6.38444099079169	-3.95368351937992	0.00055797571893	-38.3910614840635	-12.0930567674299	
18	IP	4.2808836684156	0.68251744425095	6.27219671009844	1.45858235903516E-006	2.87521267905382	5.68655465777739	
19	MR	0.03388456270511	0.02464994614108	1.37463029376118	0.18144331720092	-0.01688295169575	0.08465207710597	
20	DH	0.00122449759115	0.00413475224001	0.29614775446424	0.76956340741969	-0.00729118405324	0.00974017923553	
21								<u> </u>
<ul> <li>Regres</li> </ul>	sion (14) Regress	sion (15) Regression (16)	Regression (17) Reg	ression (18) Regression (	19) Regression (20) Regressi	ion (21) 🕨	Sum=0	- //.

QSAR

#### **Triparametric models**



QSAR

Pagina 85

From the results the best triparametric model is that with «IP» + «MR» + «SASA»

### **Triparametric models**

0.301 0.602 0.6989 0.6989 1.301 1.3979 1.4771 1.602 1.6989 1.8808 2 2.0791 2.0791 2.0791 2.176 2.2041 2.301 2.3424	Res -0 0.4 -0.1	3-10-10-10-10-10-10-10-10-10-10-10-10-10-	əl 569984562 601620200 - -	Lev 292 0.12	erages 59906563 56607010	393278	Intern	ally stu	10000000000000000000000000000000000000	Exter	mally :	studen -1.	tized p- 1791 2- 3030 5- 176 7- 125 8- 18 9- 97 2- 78 4- 68 12 88 8- 11 39 33 63 40 2- 012 0
0.301 0.602 0.6989 0.6989 1.301 1.301 1.3979 1.4771 1.602 1.6989 1.8808 2 2.0791 2.0791 2.0791 2.176 2.2041 2.301 2.3424		3-0 2-0	569984562	292 0.13	59906563	393278	•		-1.1752	•••	•	-1.	1791         24           5030         5:           76         71           25         84           18         97           97         2!           78         41           68         12           88         82           85         57           111         39           33         63           40         28           012         0
0.602 0.6989 0.6989 1.301 1.3979 1.4771 1.602 1.6989 1.8808 2 2 2.0791 2.0791 2.176 2.2041 2.301 2.3424	0.0 0.0 -0.1 -0.0 -0.0 -0.1 -	0-4- 3- 2-	-	84 0.2	56607013	201025	•		0. 5410	•••	•	•	5030         5           76         7           25         84           18         9           97         2           78         44           68         12           88         82           85         57           111         32           333         63           40         28           012         0
0.6989 0.6989 1.301 1.301 1.3979 1.4771 1.602 1.6989 1.8808 2 2 2.0791 2.0791 2.0791 2.176 2.2041 2.301 2.3424	0. -0. -0. -0. -0. -0. -0. -0. -	3 - 2 -	-				•		, • ,	•••	•	•	76 7/ 25 8/ 18 9 97 2/ 78 4/ 68 1 88 82 85 57 11 39 33 62 40 28 612 0
0.6989 1.301 1.3979 1.4771 1.602 1.6989 2.0791 2.0791 2.0791 2.176 2.2041 2.301 2.3424	-0. -0. -0. -0. -0. -0. -0. -0. -0. -0.		-				•		· •	•••	•		25 84 18 9 97 29 78 40 68 1 88 82 85 5 11 39 33 63 40 28 612 0
1.301 1.3979 1.4771 1.602 1.6989 1.8808 2 2.0791 2.0791 2.0791 2.176 2.2041 2.301 2.3424		2 -	-				•	•	· • .	•••	• . : •	•	18 9 97 2 78 4 68 1 88 8 85 5 11 3 33 6 40 2 612 0
1.301 1.3979 1.4771 1.602 1.6989 1.8808 2 2.0791 2.0791 2.176 2.2041 2.301 2.3424	0. -0. -0. -0. -0. -0. -0. -0. -	3-	-				•		· • .	•••	• •		97 2 78 4 68 1 88 8 85 5 11 39 33 6 40 28 612 (
1.3979 1.4771 1.602 1.6989 1.8808 2 2.0791 2.0791 2.176 2.2041 2.301 2.3424	-0. -0. -0. -0. -0. -0. -0. -0. 0. -0. 0. -0.	3-	-				•		· • .	•••	•		78 4 68 1 88 8 85 5 11 39 33 63 40 28 612 (
1.4771 1.602 1.6989 1.8808 2 2.0791 2.0791 2.176 2.2041 2.301 2.3424	-0. -0. -0. -0. -0. -0. 0. -0. 0. -0. 0. -0. 0. -0. -	3-	-				•	•	•	•••	:		68 1 88 8 85 5 11 39 33 63 40 28 012 (
1.602 1.6989 1.8808 2.0791 2.0791 2.176 2.2041 2.301 2.3424	-0. -0. -0. -0. -0. -0. -0. 0. -0. 0. -0. 0. -0. -	2-02-	-				•	•	•	•••	•		88 8 85 5 11 39 33 63 40 28 12 (
1.6989 1.8808 2 2.0791 2.0791 2.176 2.2041 2.301 2.3424	0. -0. -0. -0. -0. 0.	2-	-				•	•••••••••••••••••••••••••••••••••••••••	•	••	•		85 5 11 39 33 63 40 28 012 (
1.8808 2 2.0791 2.0791 2.176 2.2041 2.301 2.3424	-0. -0. -0. -0. -0.	2 -	-				•		•	••	•		11 3 33 6 40 28
2 2.0791 2.0791 2.176 2.2041 2.301 2.3424	-0. 0. -0. 0.	2-	_				•	• •		••			33 6. 40 28 012 (
2.0791 2.0791 2.176 2.2041 2.301 2.3424	0. -0. -0. 0.	2-	-				•	• •	•	••			40 20
2.0791 2.176 2.2041 2.301 2.3424	-0. -0 0.	2- 2	-				•	•					812 (
2.176 2.2041 2.301 2.3424	-0 0.	Ĩ						-					
2.2041 2.301 2.3424	0.	2						•					75 1
2.301 2.3424	0.					•	•						59 5
2.3424		d					•	•					42 8
	0.	1			•	•							64 6
2.4771	-0.	¢ ,											52 8
2.574	-0.4	4 -	1										90 2
2.602	0.4	4		• •									87 10
2.8129	-0.	¢	•										14 80
2.903	-0.	4											00 49
2.903	0.	¢		•									05 84
3	-0.	0-		1								_	89 6
3.3424	0.	1	Ó (	.5	1	1.5	0	2	2.5		3	3.5	03 7
3.4771	0.	4588	80916114	26 0.0	72860855	508795			1.2971			1.3	3125 20
3.875	0.	5250	077934532	294 0.1	19864088	362865			1.5235			1.	5577 1
3.875	0.	6724	480428845	581 0.10	64666340	031334			2.0029			2.	1088 4
	2.8129 2.903 2.903 3.3424 3.4771 3.875 3.875	2.902 0. 2.8129 -0. 2.903 -0. 2.903 0. 3 -0. 3.3424 0. 3.4771 0. 3.875 0. 3.875 0.	2.002 0.3 2.8129 -0.0 2.903 -0.2 2.903 0.0 3 -0.2 0.1 3.3424 0.1 3.4771 0.458 3.875 0.525 3.875 0.672	2.8129 -0. 2.903 -0. 2.903 -0. 2.903 -0. 3.3424 0. 0. 3.4771 0.458809161140 3.875 0.525077934532 3.875 0.672480428845	2.8129 -0.0 2.903 -0.1 2.903 -0.1 2.903 0.0 3.3424 0.1 0 0.5 3.4771 0.45880916114026 0.0 3.875 0.52507793453294 0.1 3.875 0.67248042884581 0.1	2.8129 -0.6 2.903 -0.2 2.903 -0.2 2.903 0.6 3.3424 0.1 0.5 1 3.3471 0.45880916114026 0.072860855 3.875 0.52507793453294 0.119864084 3.875 0.67248042884581 0.164666346	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	2.8129 -0.0 2.903 -0.1 2.903 0.0 3.3424 0.0 3.34771 0.45880916114026 0.07286085508795 3.875 0.52507793453294 0.11986408862865 3.875 0.67248042884581 0.16466634031334	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	2.802       0.3         2.8129       0.0         2.903       0.1         3.90.1       0         3.3424       0.1         0.45880916114026       0.07286085508795         3.875       0.52507793453294         0.1586408862865       1.5235         1.15       2.0029         2.15       3.875         0.52507793453294       0.11986408862865         1.5235       1.5235         3.875       0.67248042884581         0.16466634031334       2.0029

#### The fitting plot can be made by the «Insert $\rightarrow$ Chart» menu

#### Compare result with those published

 $^{SUCROSE}$ PS1 = 5.08109 × IP – 27.7204.

 $r^2=0.727709,\; rCV^2=0.713409,\; SE=0.1177,\; SEE=0.5028,\; t\text{-value}=8.4963, P\text{-value}=0,\; DOF=0.7177,\; n=29.$ 

 $\label{eq:sucrosseps2} \begin{array}{l} {}^{SUCROSE}PS2 = 4.07088 \times IP + 0.616562 \times \omega - 23.1215. \\ r^2 = 0.842159, \ rCV^2 = 0.805992, \ SE = 0.0833, \ SEE = 0.3829, \\ t\text{-value} = 12.0028, \ P\text{-value} = 0, \ DOF = 0.8363, \ n = 29. \end{array}$ 

 $^{SUCROSE}{\rm PS3} = 4.63976 \times {\rm IP} + 0.138893 \times {\rm MR} - 0.0712653 \times {\rm SASA} - 25.5542.$ 

 $r^2=0.865448,\; rCV^2=0.816901,\; SE=0.0759,\; SEE=0.3536,\; t\text{-value}=13.1759,\; P\text{-value}=0,\; DOF=0.8604,\; n=29.$ 

Are the same!

QSAR

Pagina 87

Comparing the results with those reported in the publication the same results were obtained using data not normalized. The same results are obtained either with normalized or autoscaled data!