قسم الهندسة الكيمياوية

وزارة التعليم العالي والبحث العلمي
الجامعة التكنولوجية

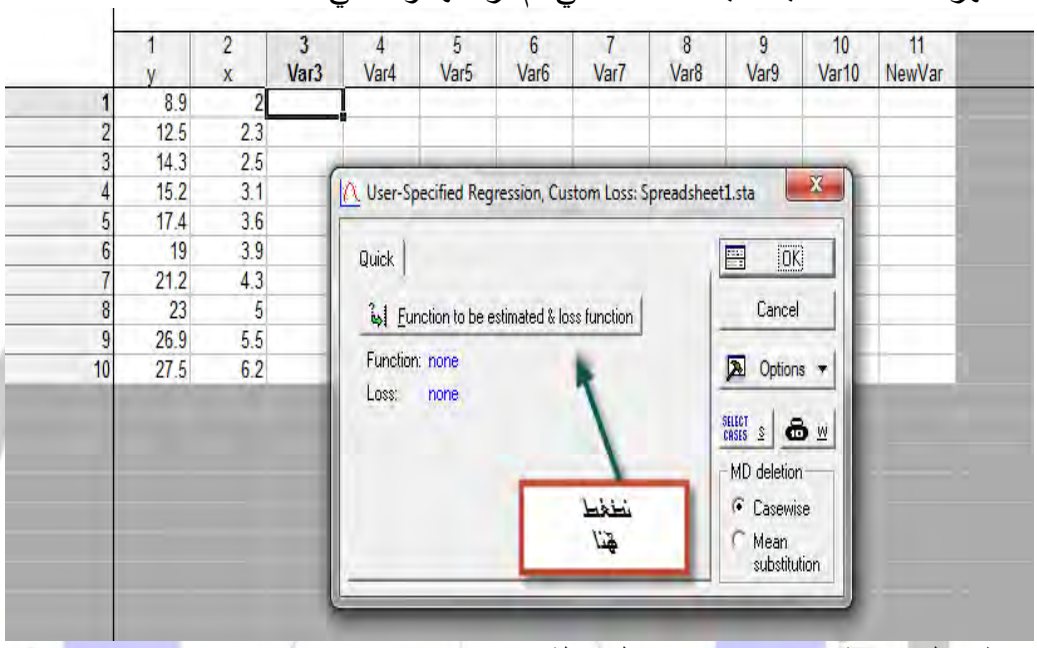
Statistics Laboratory Guide

Statistics Laboratory Guide

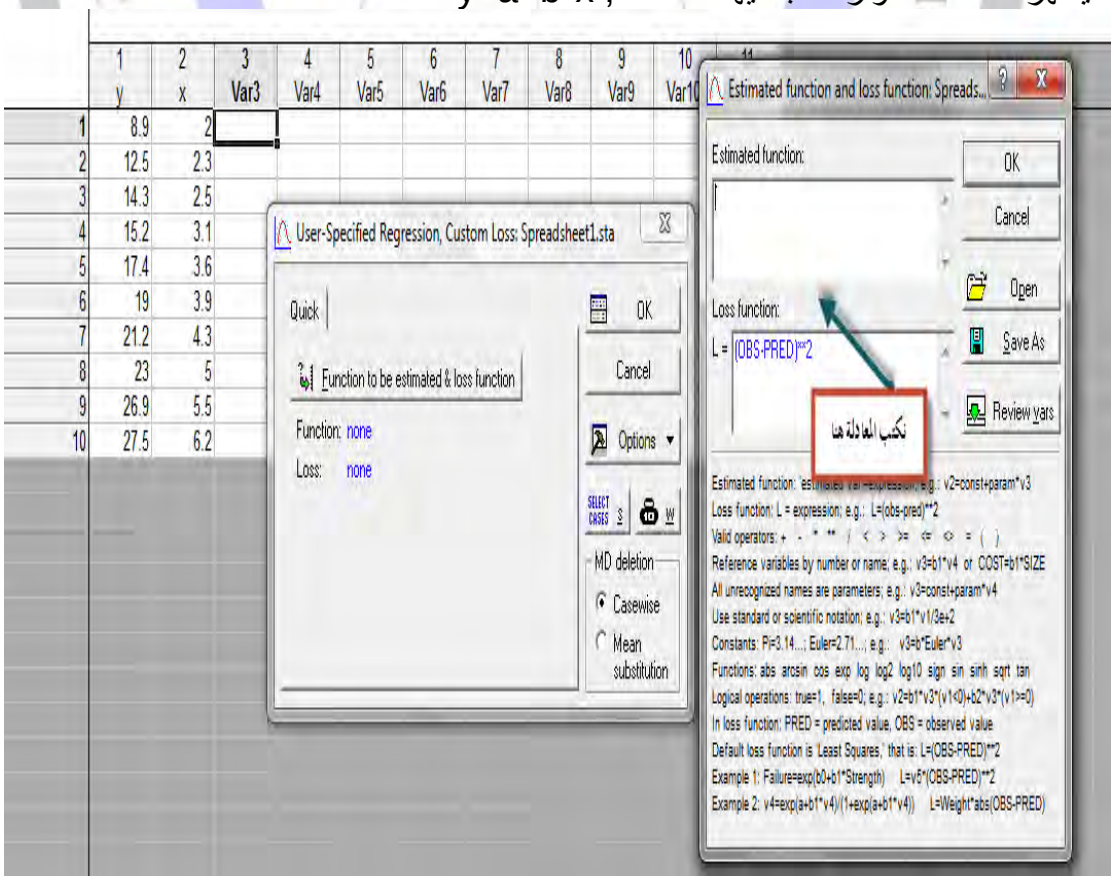
Prepared by:

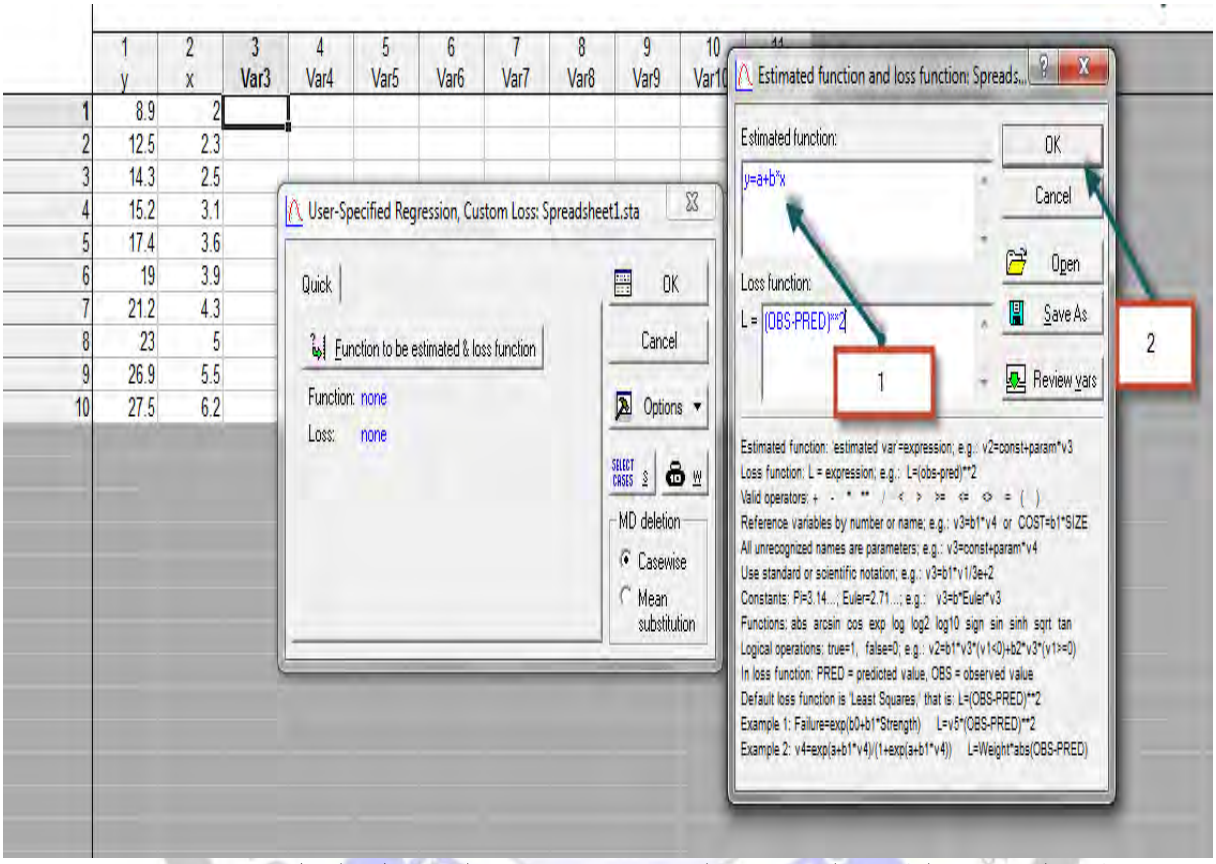
Lecturer: Mahir A. Abdul Rahman

ستظهر لنا نافذة تطلب كتابة المعادلة التي تم فرضها وكالأتي :

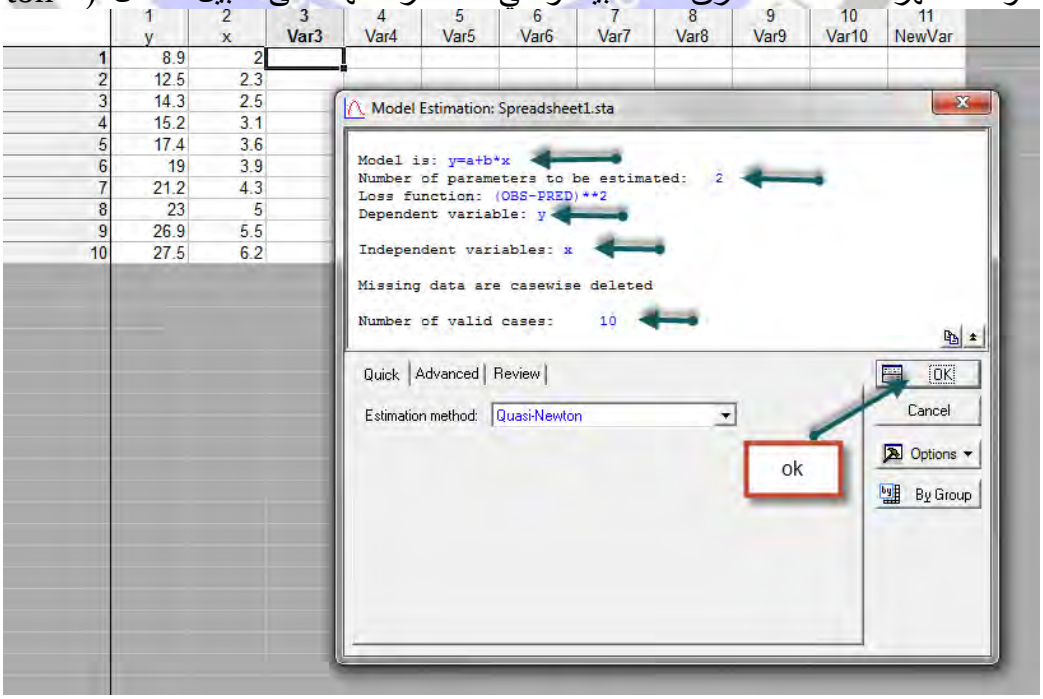


سيظهر لنا شاشة حوار نكتب فيها المعادلة , y=a+b*x

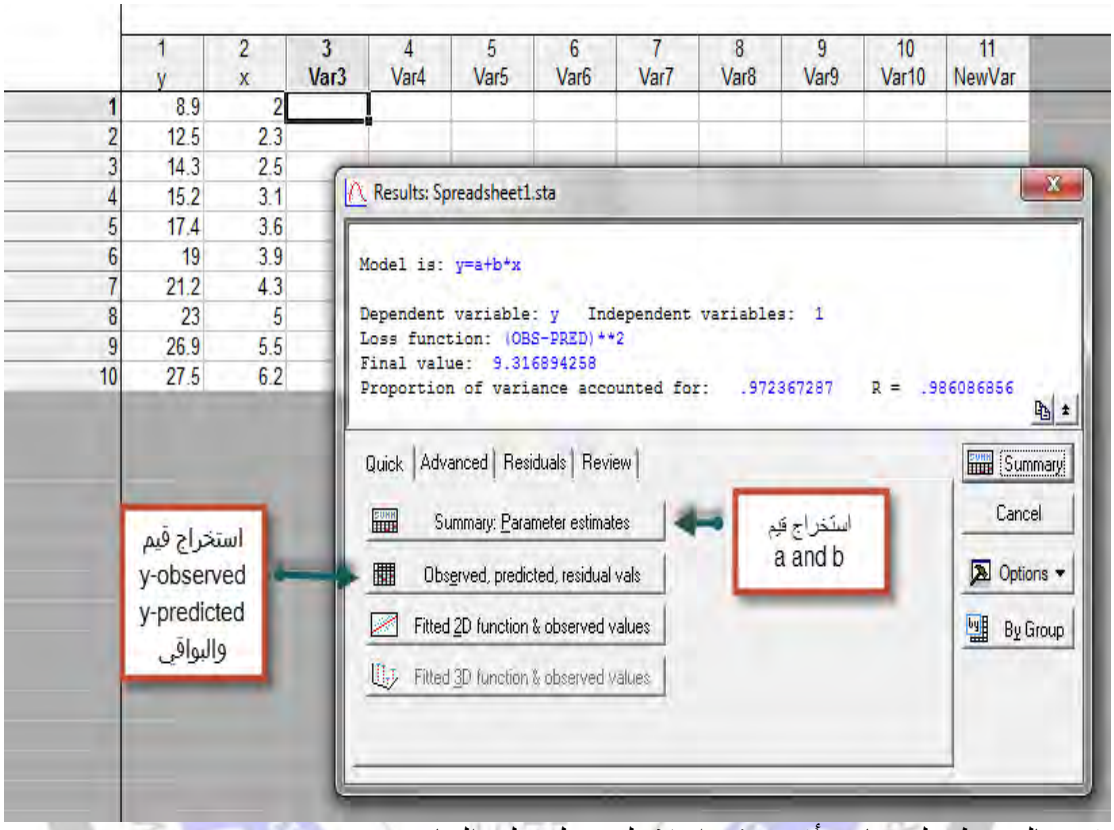




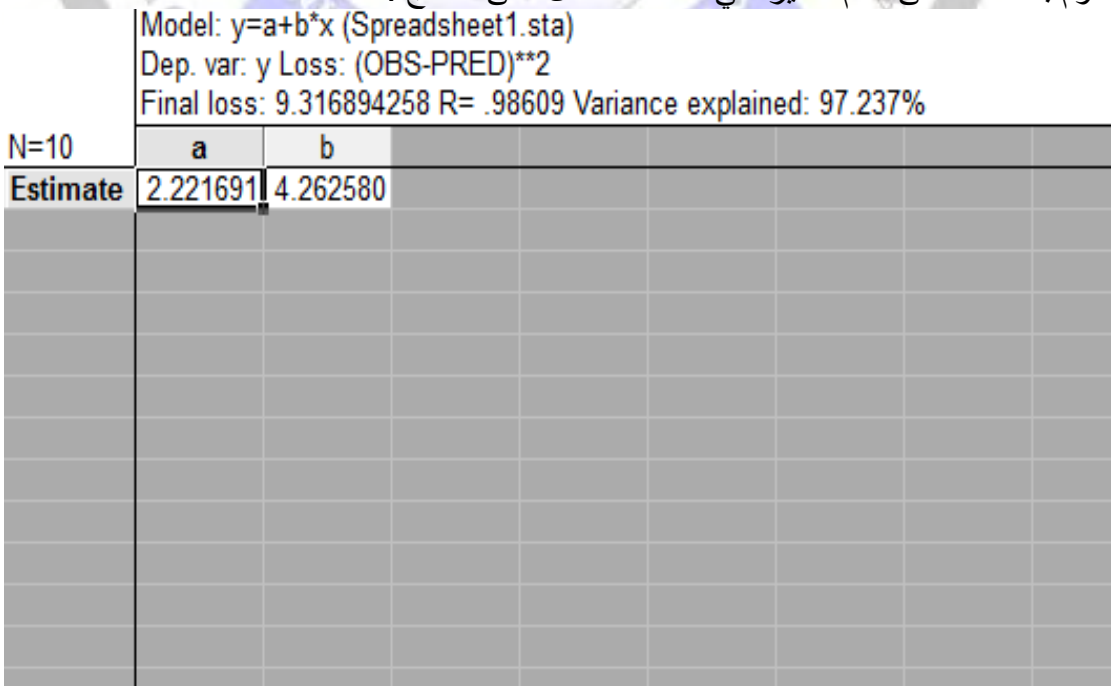
سوف تظهر لنا نافذة الطرق الحسابية والتي سنختار منها على سبيل المثال (Quasi-Newton)



ونضغط (OK) سيقوم البرنامج بحساب القيم التخمينية وظهور شاشة الحوار التالية :



نقوم بالضغط على ماتم تأشيرة في اعلاة لنحصل على النتائج :



Model is: y=a+b*x (Spreadsheet1.sta)
Dep. Var. : y

	Observed	Predicted	Residuals
1	8.90000	10.74685	-1.84685
2	12.50000	12.02563	0.47437
3	14.30000	12.87814	1.42186
4	15.20000	15.43569	-0.23569
5	17.40000	17.56698	-0.16698
6	19.00000	18.84575	0.15425
7	21.20000	20.55079	0.64921
8	23.00000	23.53459	-0.53459
9	26.90000	25.66588	1.23412
10	27.50000	28.64969	-1.14969

2- نقوم بنفس الخطوات السابقة للحصول على نتائج المعادلة الثانية

Y=a*x

Model: Y=a*x (Spreadsheet1.sta)
Dep. var: y Loss: (OBS-PRED)**2
Final loss: 14.698381269 R= .97796 Variance explained: 95.641%

N=10	a
Estimate	4.778066

Model is: Y=a*x (Spreadsheet1.sta)
Dep. Var. : y

	Observed	Predicted	Residuals
1	8.90000	9.55613	-0.65613
2	12.50000	10.98955	1.51045
3	14.30000	11.94517	2.35483
4	15.20000	14.81201	0.38799
5	17.40000	17.20104	0.19896
6	19.00000	18.63446	0.36554
7	21.20000	20.54569	0.65431
8	23.00000	23.89033	-0.89033
9	26.90000	26.27937	0.62063
10	27.50000	29.62401	-2.12401

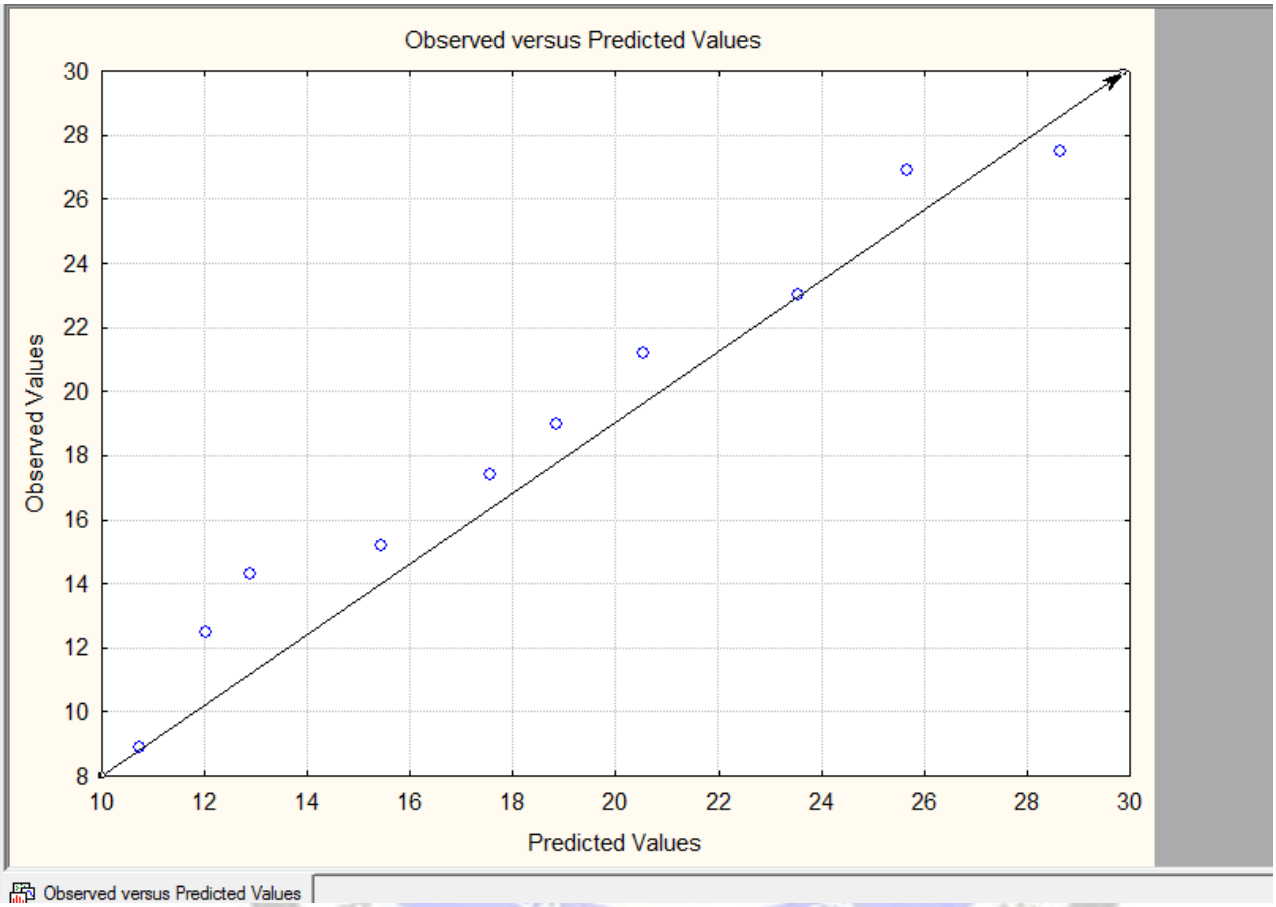
WAYS TO VALIDATE THE EQUATION IMPOSED

وتوجد ثلاث طرق :

Line (45) -1

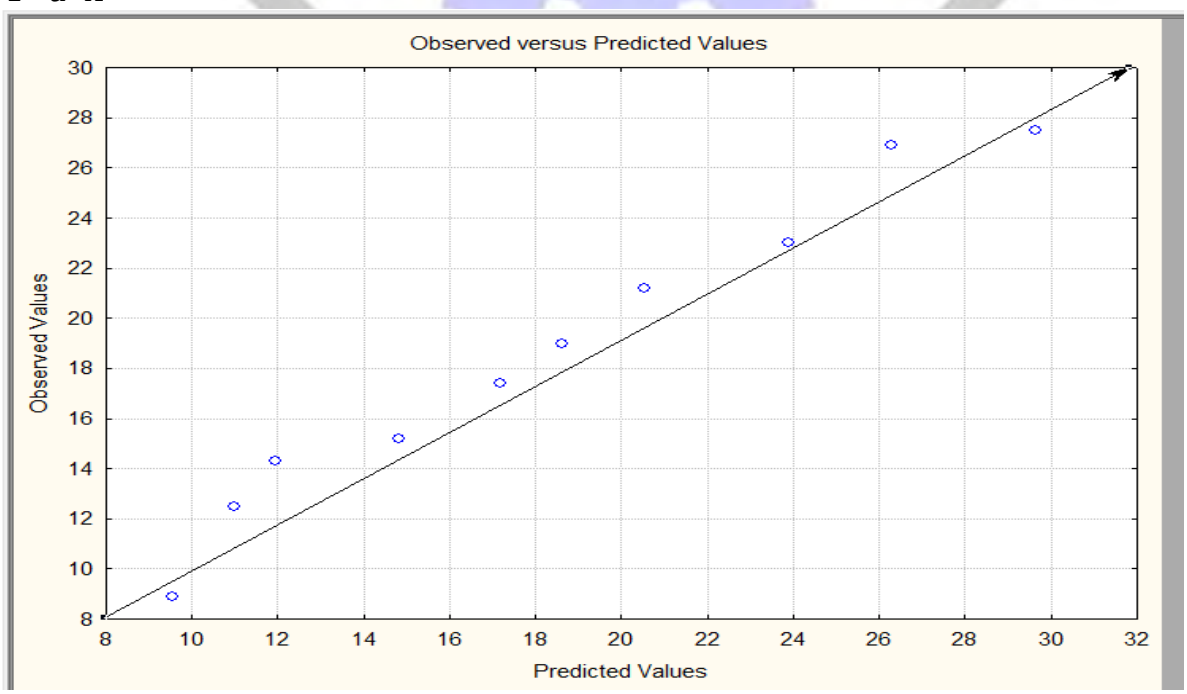
الضغط على Residual تظهر لنا عدة خيارات نختار منها Predicted Vs.Observed.val
سنحصل على شكل بياني يضم قيم كلا" من (y-observed) و (y- predicted) فكلما كانت القيم واقعة على خط (45) درجة او قريبة منة , يعني ذلك ان المعادلة المفروضة صحيحة وكلما ابتعدت عنة يدل على ان المعادلة المفروضة غير صحيحة.
وبالتطبيق على المثال السابق نجد ان

Y=a+b*x



بينما المعادلة الثانية

Y=a*x



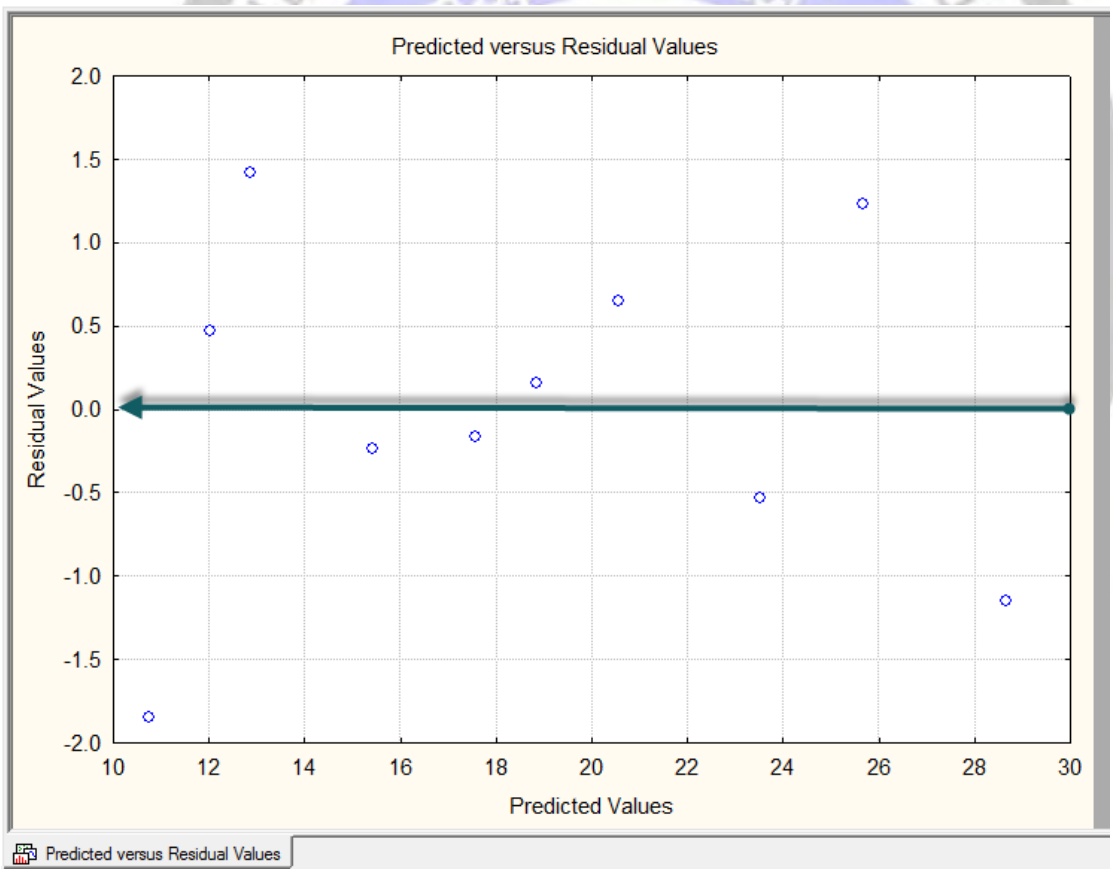
يتضح مما سبق ان المعادلة الأولى Y=a+b*x هي الاصح لوقوع ولقرب النقاط من خط (45) درجة اذا ماقورنت بالمعادلة الثانية Y=a*x

-2 (Residual Plot)

بالضغط على Residual تظهر لنا عدة خيارات نختار منها Predicted Vs. Residual. Val للحصول على الشكل البياني التالي والذي يوضح العلاقة بين قيم indep.x (y-predicted) و (Residual . val) , فكلما اقتربت النقاط من (x-axis) يعني ان المعادلة المفروضة صحيحة والعكس صحيح .

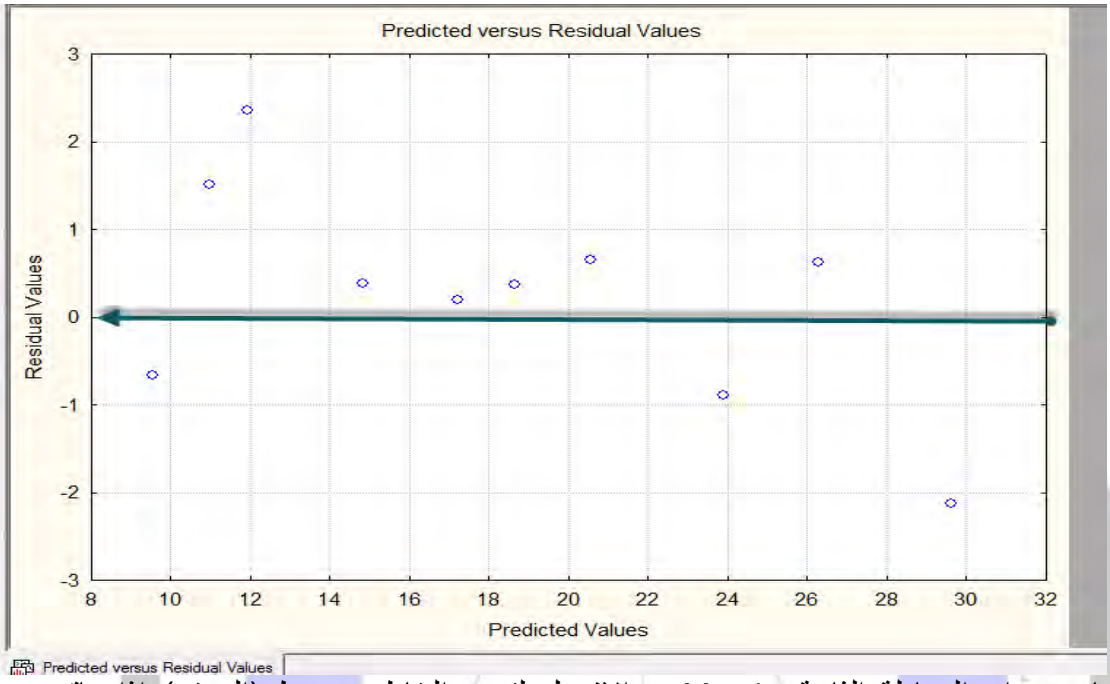
ومن خلال التطبيق نجد ان :

Y=a+b*x



بينما نجد ان الشكل البياني للمعادلة الثانية كان

Y=a*x



يتضح مما سبق ان المعادلة الثانية Y=a*xهي الافضل لقرب النقاط من خط (الصفر) اذا ماقورنت بالأولى Y=a+b*x

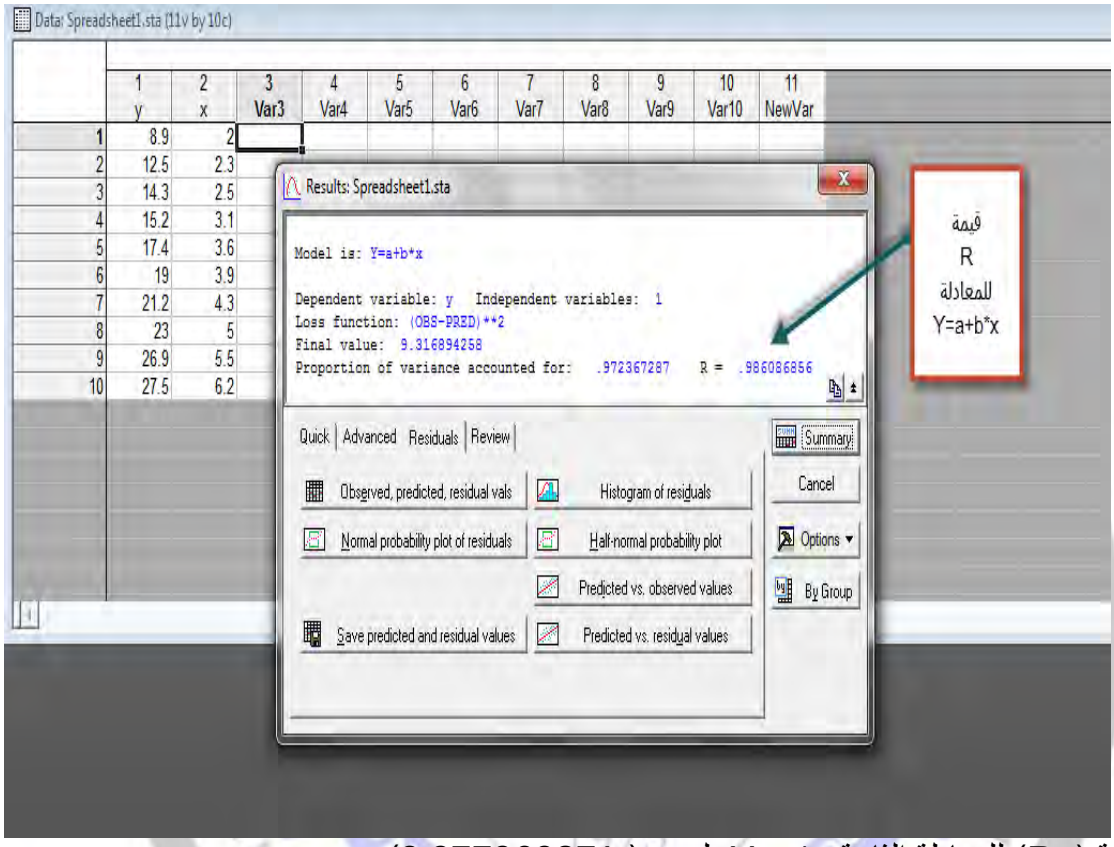
Correlation Coefficient

والذي نرمز لة بالرمز (R) ويعتبرمن الطرق البسيطة والدقيقة , وهناك عدة حالات :

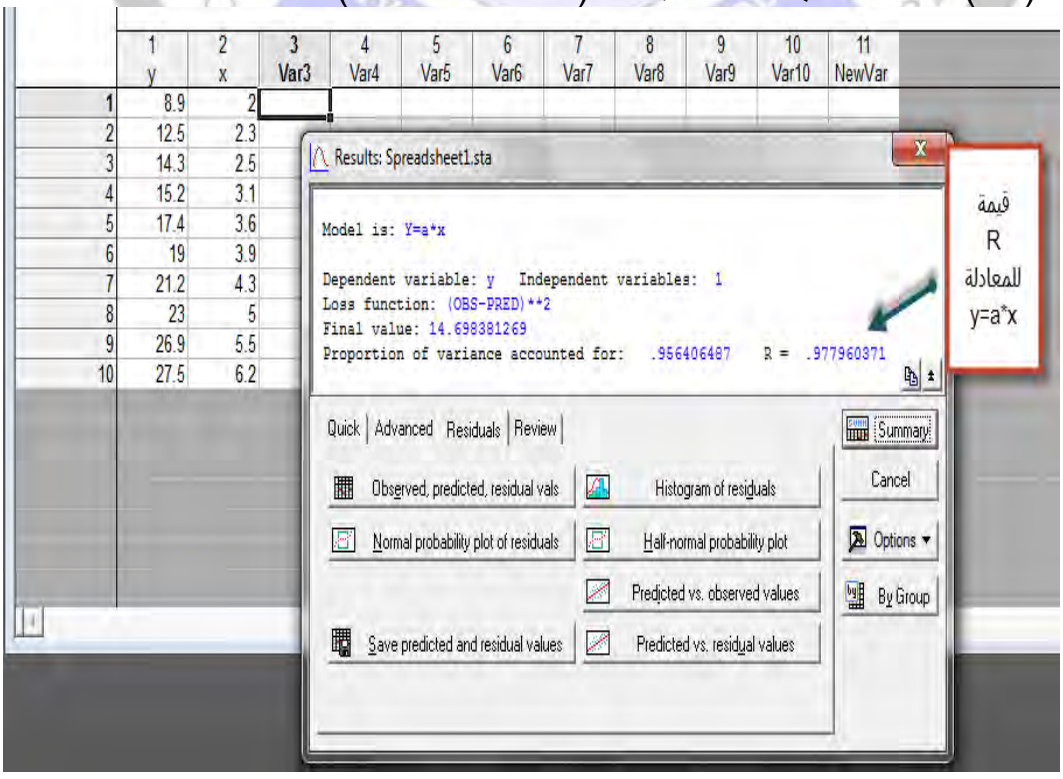
- عندما تكون النقاط في الشكل البياني بين (Y-obs.Vs. Y-pre) متطابقة والميل موجب عندها تكون قيمة (R=1) وهي من الحالات النادرة حيث انة لايمكن ان نحصل على معادلة تتطابق فيها كل (Data) .
- عندما تكون النقاط تقريبا" على استقامة واحدة , أي وجود بعض النقاط تنحرف عن الخط المستقيم بالإضافة الى كون الميل موجب , هنا تتراوح قيمة (R=0.9-0.99) وهذا يعني ان المعادلة المفروضة تنطبق على معظم النقاط , بمعنى اخر كلما اقتربت المعادلة من الواحد الصحيح كانت المعادلة ادق .
- عندما تقترب قيم (R) من الصفر يعني ان النقاط مبعثرة ولايمكن ان تتحدد باي علاقة.

ومن خلال التطبيق نلاحظ ان قيمة (R=0.986086856) للمعادلة الأولى

Y=a+b*x



بينما قيمة (R) للمعادلة الثانية Y=a*x بلغت (0.977960371)



وعلية قيمة (R) للمعادلة الأولى Y=a+b*x هي الأقرب للدقة من المعادلة الثانية Y=a*x لكون قيمة اقرب للواحد صحيح .

سؤال :

اذا علمت ان قيم (x,y) هي

X	4.5	11.2	16.6	20.4	24.2
y	100	200	300	400	500

1- باستخدام y=a*x^0.4

2- Y=x^2+a*x+b

الحل

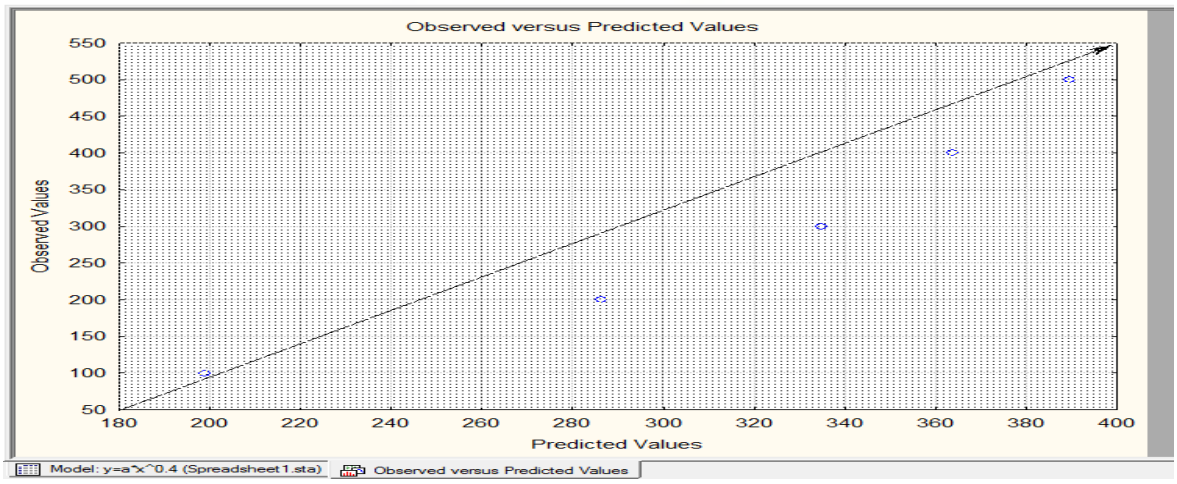
المعادلة الاولى y=a*x^0.4

a=108.93189

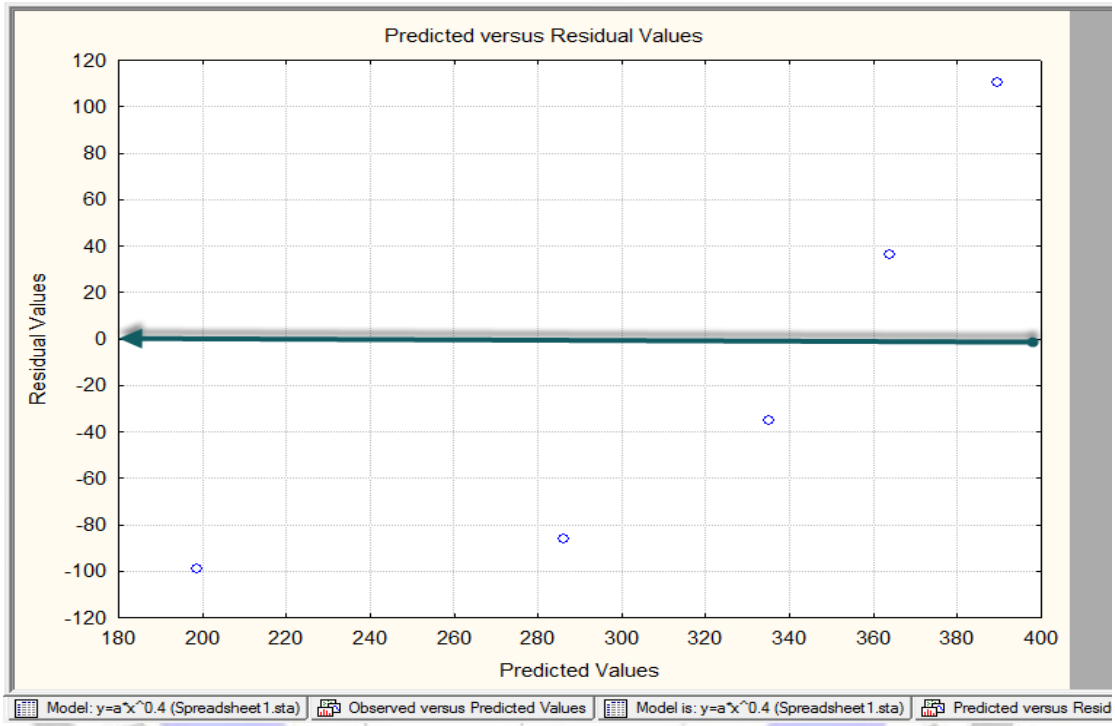
Model is: y=a*x^0.4 (Spreadsheet1.sta)
Dep. Var. : y

	Observed	Predicted	Residuals
1	100.0000	198.8108	-98.8108
2	200.0000	286.3137	-86.3137
3	300.0000	335.1183	-35.1183
4	400.0000	363.9209	36.0791
5	500.0000	389.6557	110.3443

A- Line (45)



B- **Residual Plot**



C- Correlation Coefficient

R = .825078903

الم ع دلة الث انية

Y=x^2+a*x+b

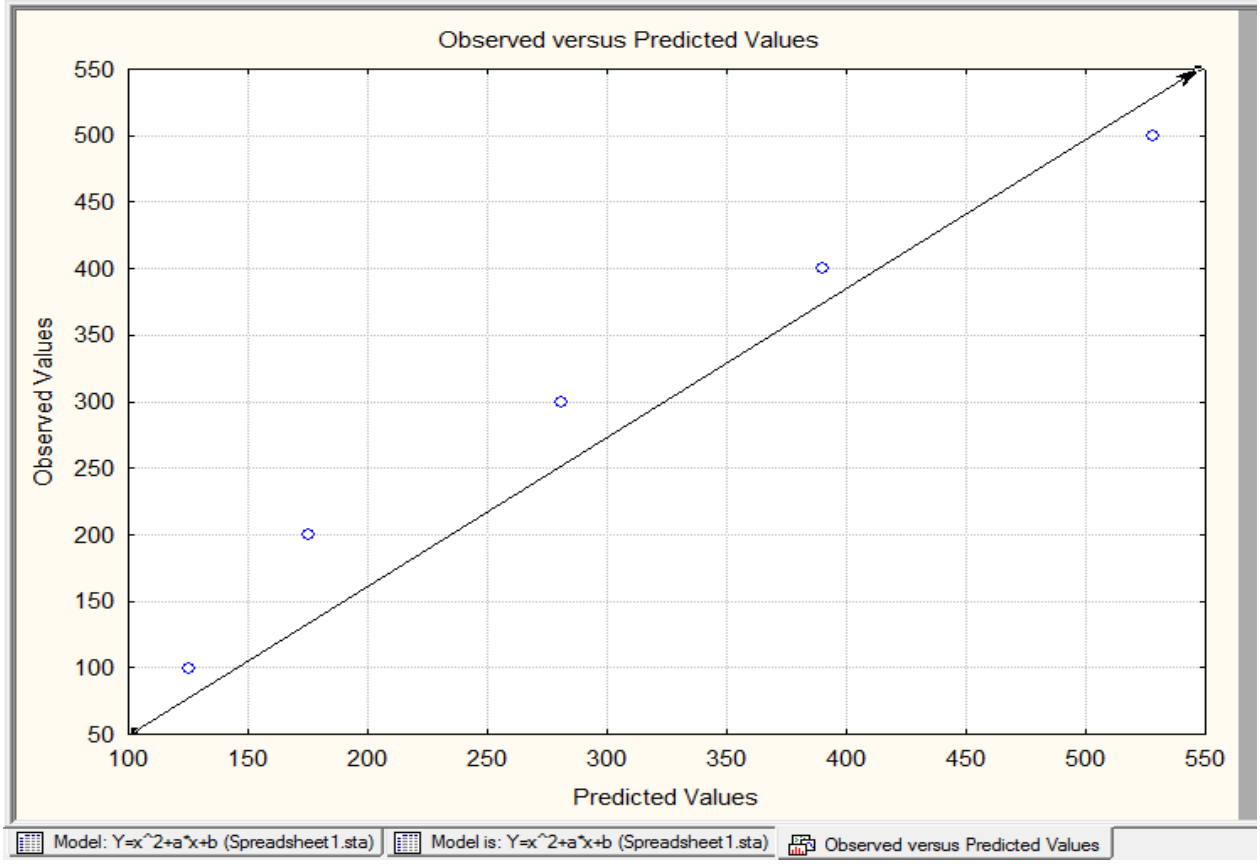
Model: Y=x^2+a*x+b (Spreadsheet1.sta)
Dep. var: y Loss: (OBS-PRED)**2
Final loss: 2516.0759706 R= .98734 Variance explained: 97.484%

N=5	a	b
Estimate	-8.26367	142.4853

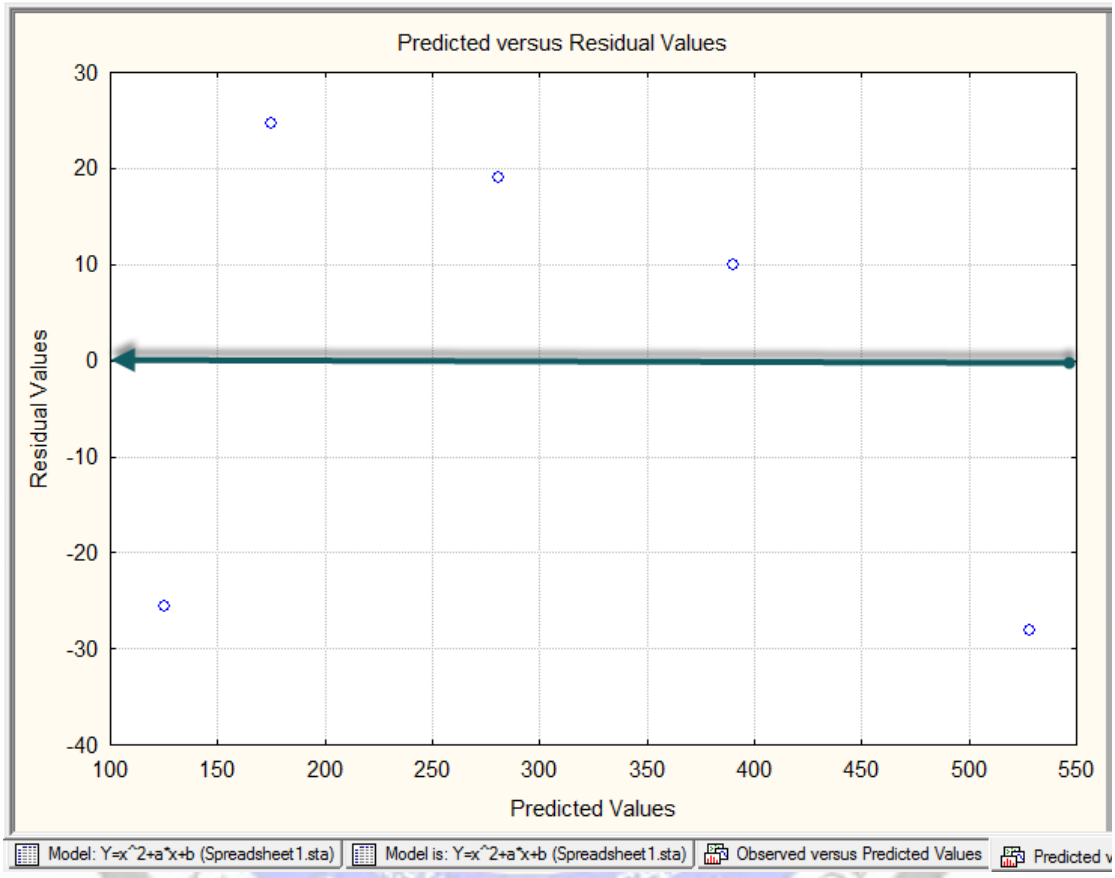
Model is: Y=x^2+a*x+b (Spreadsheet1.sta)
Dep. Var. : y

	Observed	Predicted	Residuals
1	100.0000	125.5488	-25.5488
2	200.0000	175.3722	24.6278
3	300.0000	280.8683	19.1317
4	400.0000	390.0664	9.9336
5	500.0000	528.1444	-28.1444

A- Line (45)



B- **Residual Plot**



C- Correlation Coefficient:

R = .987339476

وعلية المعادلة Y=x^2+a*x+b هي الادق والاصح من معادلة y=a*x^0.4

الانحدار الخطي المتعدد Multiple Linear Regression

يعد الإنحدار الخطي المتعدد من الأساليب الإحصائية المتقدمة والتي تضمن دقة الإستدلال من أجل تحسين نتائج البحث عن طريق الإستخدام الأمثل للبيانات في إيجاد علاقات سببية بين الظواهر موضوع البحث .

والإنحدار الخطي المتعدد هو عبارة عن إيجاد معادلة رياضية تعبر عن العلاقة بين متغيرين وتستعمل لتقدير قيم سابقة ولتنبؤ قيم مستقبلية ، وهو عبارة أيضاً عن إنحدار للمتغير التابع (Y) على العديد من المتغيرات المستقلة X1 , X2 ,...XK لذا فهو يستخدم في التنبؤ بتغيرات المتغير التابع الذي يؤثر فيه عدة متغيرات مستقلة أي تعتمد فكرته على العلاقات الدلالية التي تستخدم ما يعرف بشكل التشتت أو الانتشار ، فبإمكاننا التنبؤ بالمستوى الرقمي في فعالية رمي المطرقة على سبيل المثال إعتماداً على دراسة حالات أخرى للرامي كالعمر الزمني والعمر التدريبي والمهارة والمواصفات الجسمية وغيرها .

إن الإنحدار الخطي المتعدد ليس مجرد أسلوب واحد وإنما مجموعة من الأساليب التي يمكن أستخدامها لمعرفة العلاقة بين متغير تابع مستمر وعدد من المتغيرات المستقلة التي عادةً ما تكون مستمرة)

والمعادلة الخطية في الإنحدار الخطي المتعدد هي :

$$Y = a + b1X1 + b2X2 + \ldots\ldots + e$$

حيث أن

Y = المتغير التابع

a = قيمة ثابتة Constant أو Intercept

b1 = ميل الإنحدار y على المتغير المستقل الأول

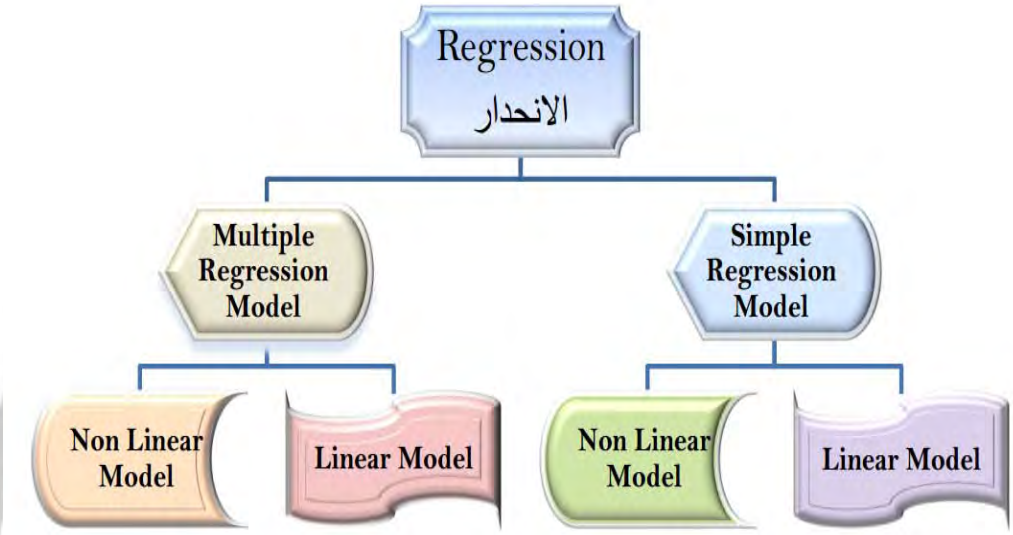
b2 = ميل الإنحدار y على المتغير المستقل الثاني

X1 = المتغير المستقل الأول

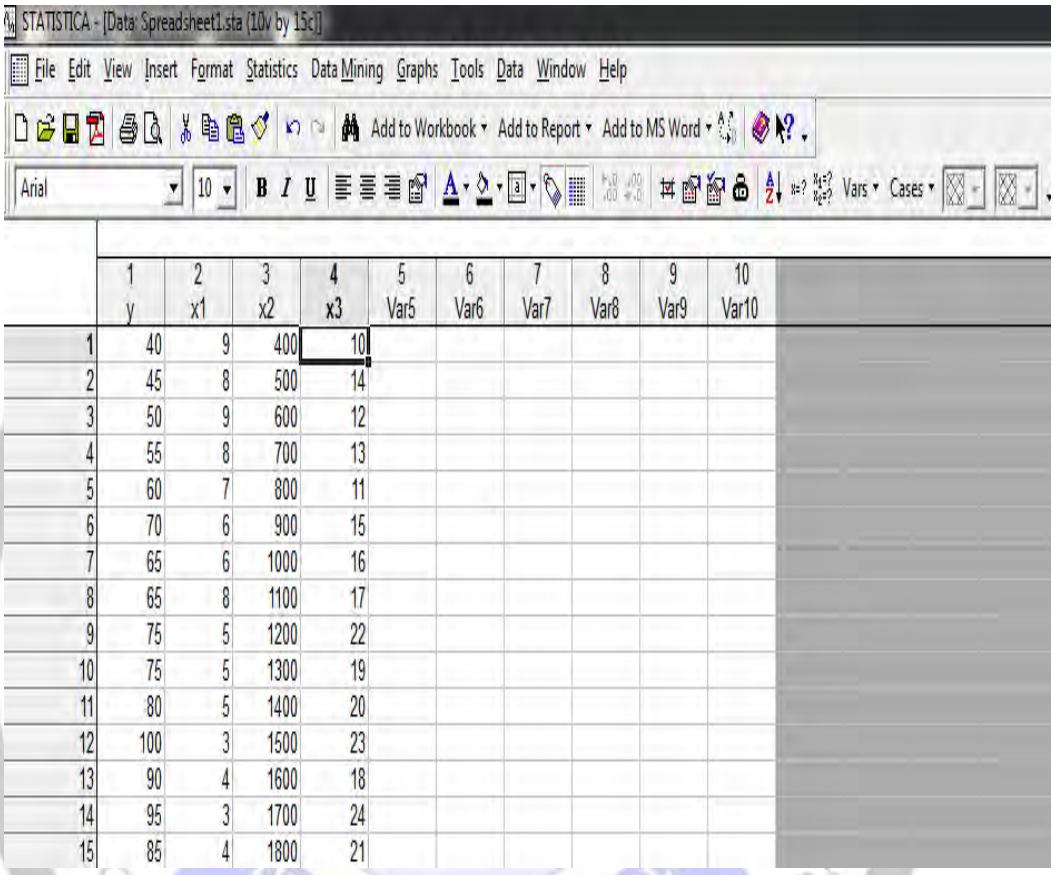
X2 = المتغير المستقل الثاني

ويمكن استخدام الإنحدار الخطي المتعدد في حالة توافر الشروط التالية :

1. أن تكون العلاقة خطية بين المتغيرات المستقلة والمتغير التابع .
2. أن تكون البيانات موزعة توزيعاً طبيعياً للمتغيرات المستقلة والمتغير التابع
3. يجب أن تكون قيم المتغير التابع من المستوى الترتيبي على الأقل .

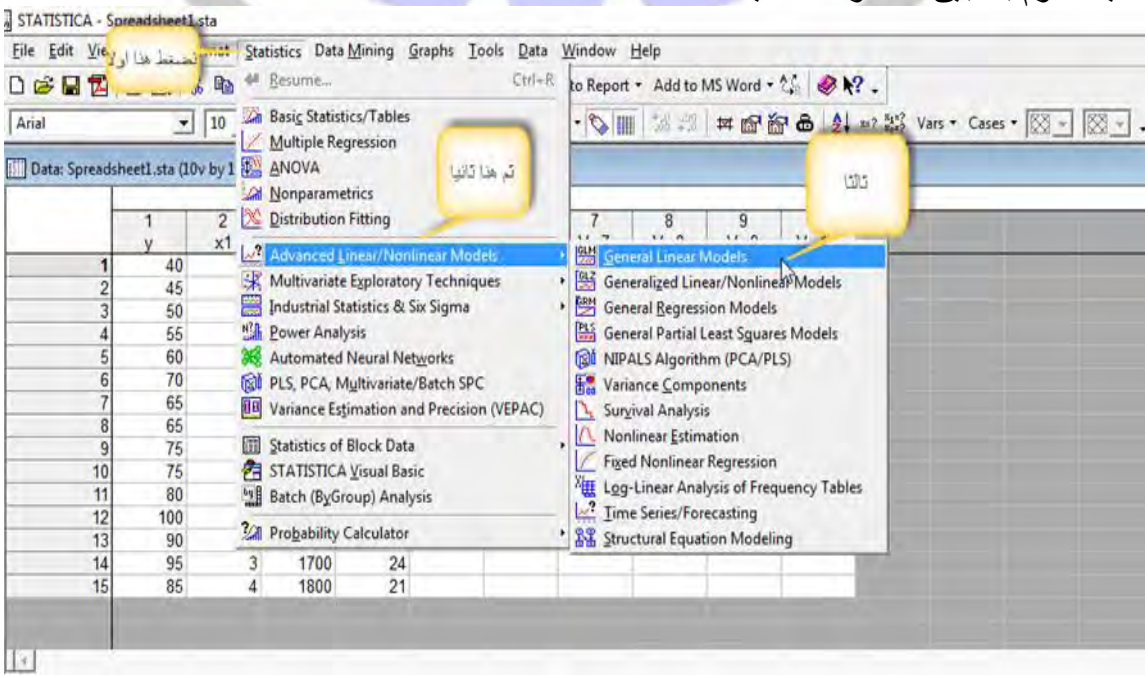


بعد الحصول على نتائج معادلة الانحدار يجب علينا أن نبين هل أن هذه المعاملات مقبولة من الناحية الإحصائية أي معنوية احصائياً مع التنويه بأن المعنوية تكون لكل معامل على حدة .
ولكي نحكم على معنوية معاملات الإنحدار نستعين باختبار ***T*** ومستوى الاحتمالية المقابل له وبالطبع فإن برنامج ***SPSS*** سيقوم تلقائياً باستخراج اختبار ***T*** ومستوى الاحتمالية المقابل له .
كما سيتم الحصول على إحصائيات تستخدم لمعرفة المعنوية الإجمالية للنموذج ومنها (R) ، (R^2) ، (R^{2-}) .
R هو معامل الأرتباط البسيط والذي يقيس قوة العلاقة بين متغيرين أو أكثر ، أما ***R^2*** فهو يسمى بمعامل التحديد والذي يستخدم لمعرفة القوة التفسيرية للنموذج المقدر (المعادلة المقدرة) في حالة الإنحدار الخطي البسيط (متغير مستقل واحد مع متغير معتمد واحد) ، أما ***R^{2-}*** فهو يستخدم لتفسير القوة التفسيرية لنموذج الانحدار الخطي المتعدد (لأنه يأخذ بنظر الاعتبار عدد المتغيرات المستقلة ولذلك يسمى بالمصحح لأنه بالأصل مشتق من ***R^2***) .
كما نستخدم أيضاً أحصائية ***F*** للحكم على معنوية النموذج المقدر ككل عند مستوى معنوية معين .
سنتناول المثال الموضح في الجدول التالي لإيجاد معادلة خط الإنحدار المتعدد للبيانات المعلقة بأختبار (15) طالباً إذ تمثل هذه البيانات العلاقة بين (أختبار الوثب العريض من الثبات ***Y***) والعوامل المؤثرة عليها وهي (أختبار الشد لأعلى على العقلة ***X1***) و (أختبار السعة الحيوية ***X2***) و (أختبار الوثب العمودي من الثبات ***X3*** وسيتم الحصول على نتائج تقدير معادلة الانحدار الخطي المتعدد وكما يلي :

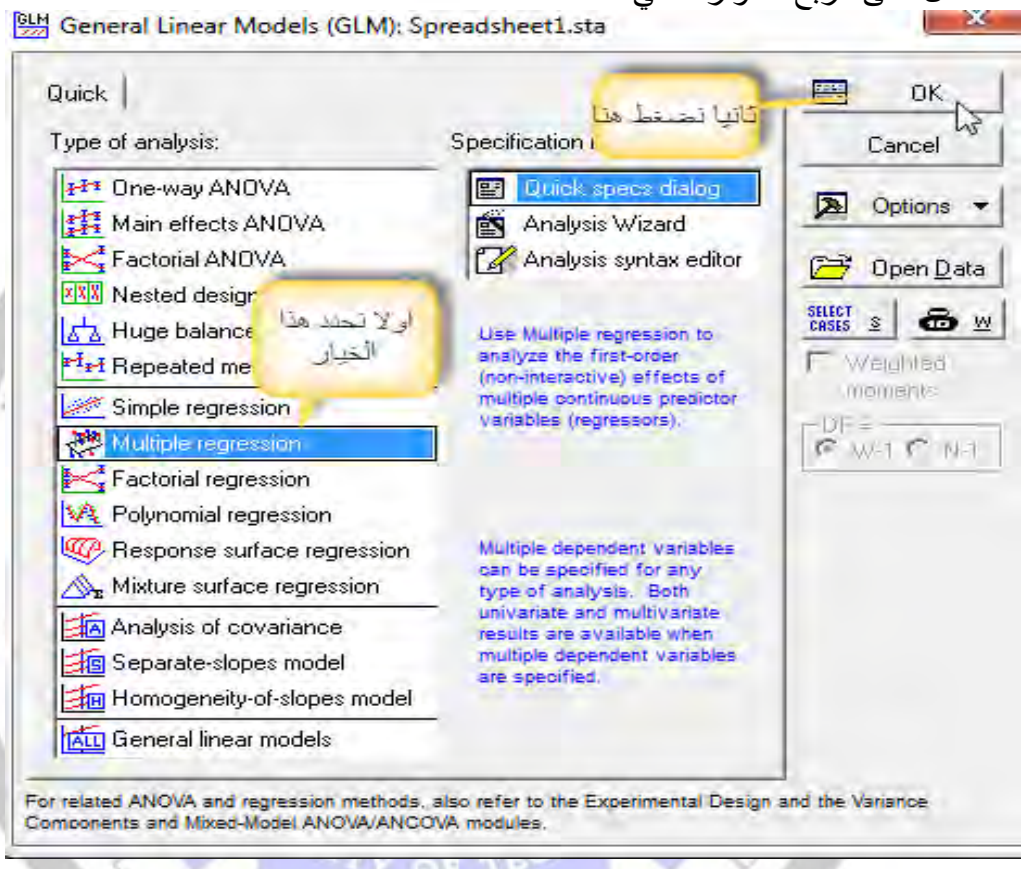


	1 y	2 x1	3 x2	4 x3	5 Var5	6 Var6	7 Var7	8 Var8	9 Var9	10 Var10
1	40	9	400	10						
2	45	8	500	14						
3	50	9	600	12						
4	55	8	700	13						
5	60	7	800	11						
6	70	6	900	15						
7	65	6	1000	16						
8	65	8	1100	17						
9	75	5	1200	22						
10	75	5	1300	19						
11	80	5	1400	20						
12	100	3	1500	23						
13	90	4	1600	18						
14	95	3	1700	24						
15	85	4	1800	21						

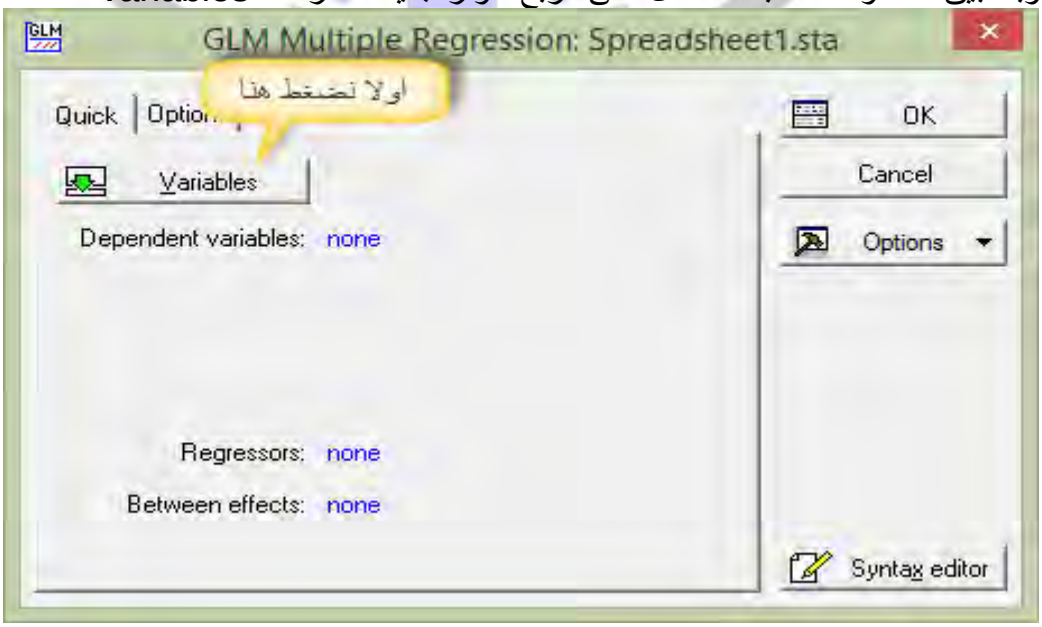
أولا / نقوم بادخال البيانات في محرر بيانات البرنامج
ثانيا / نقوم بتطبيق الخطوات التالية :



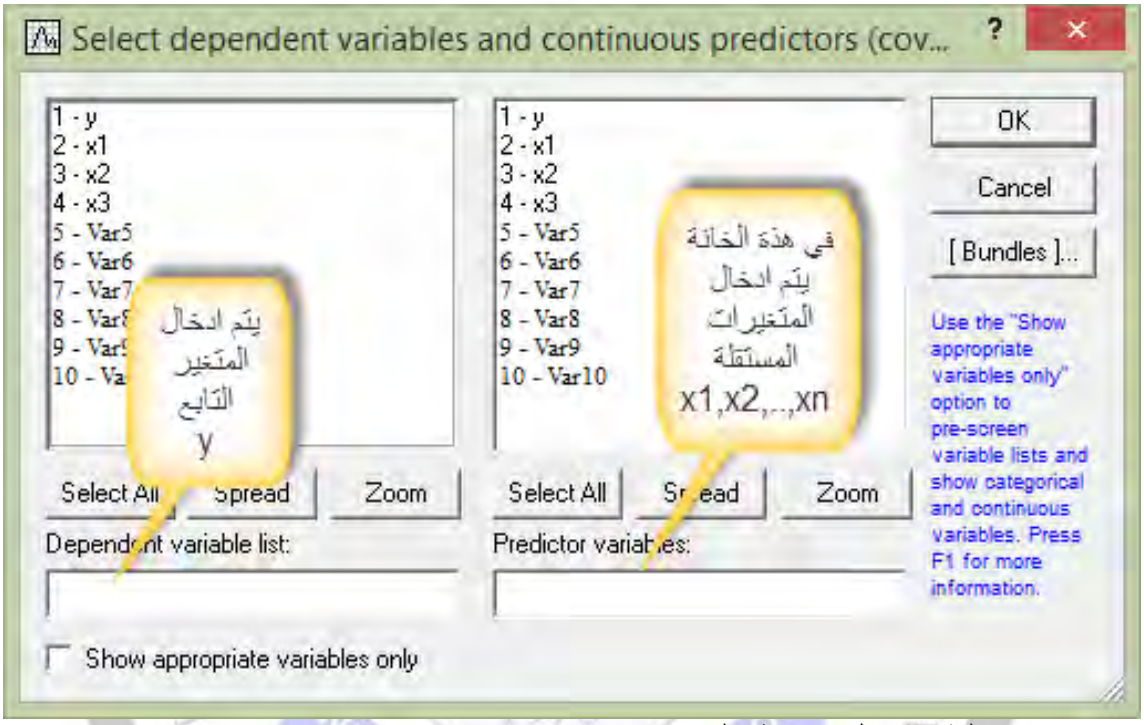
سنحصل على مربع الحوار التالي



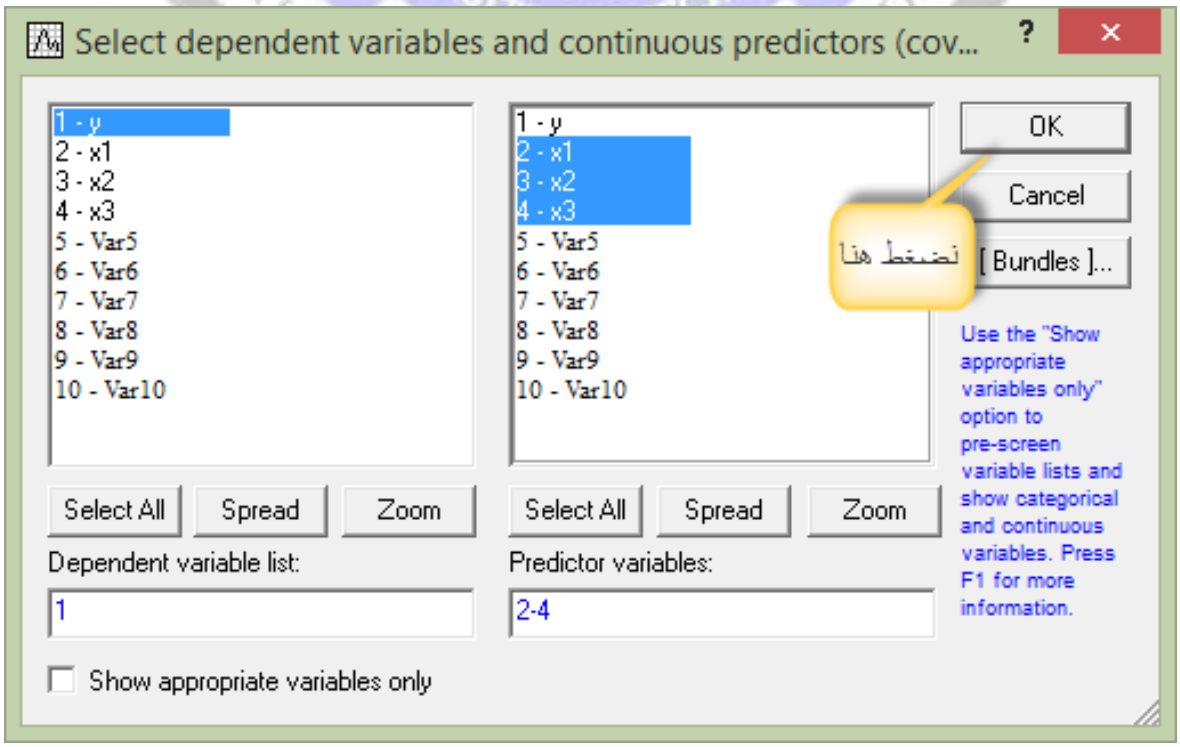
وبتطبيق الخطوات السابقة نحصل على مربع حوار جديد نختار منه variables



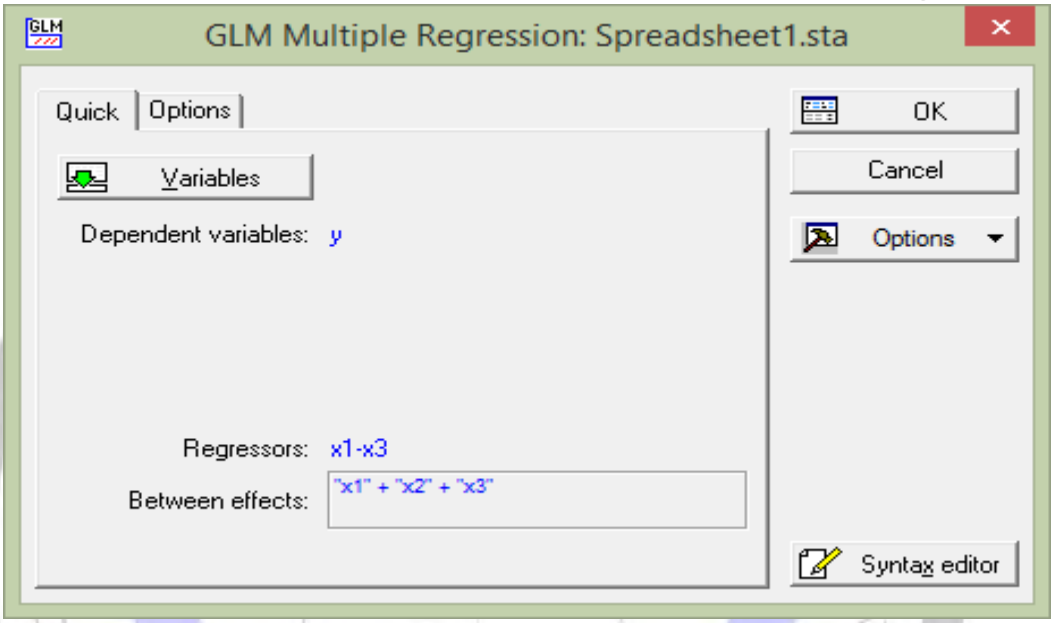
ومن ثم نخل المتغير التابع y في خانة Dependent variables بينما ندخل المتغيرات x في خانة Predictor variables :

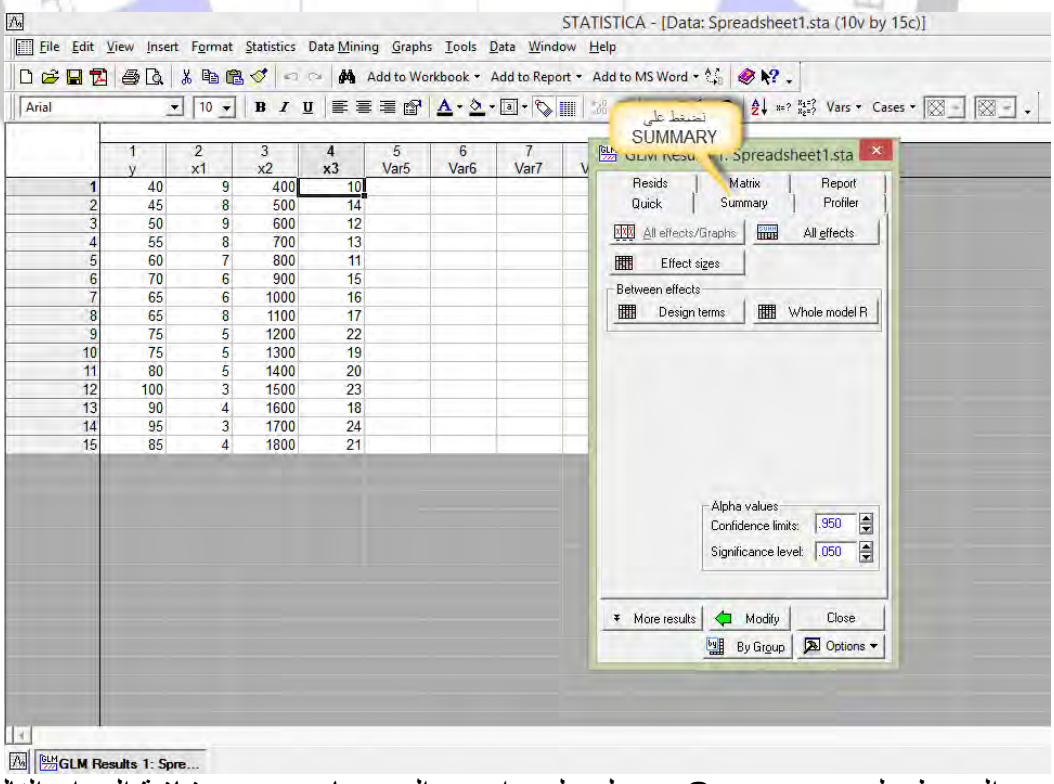


وعند اتمام عملية الادخال نضغط على OK وكما موضح :

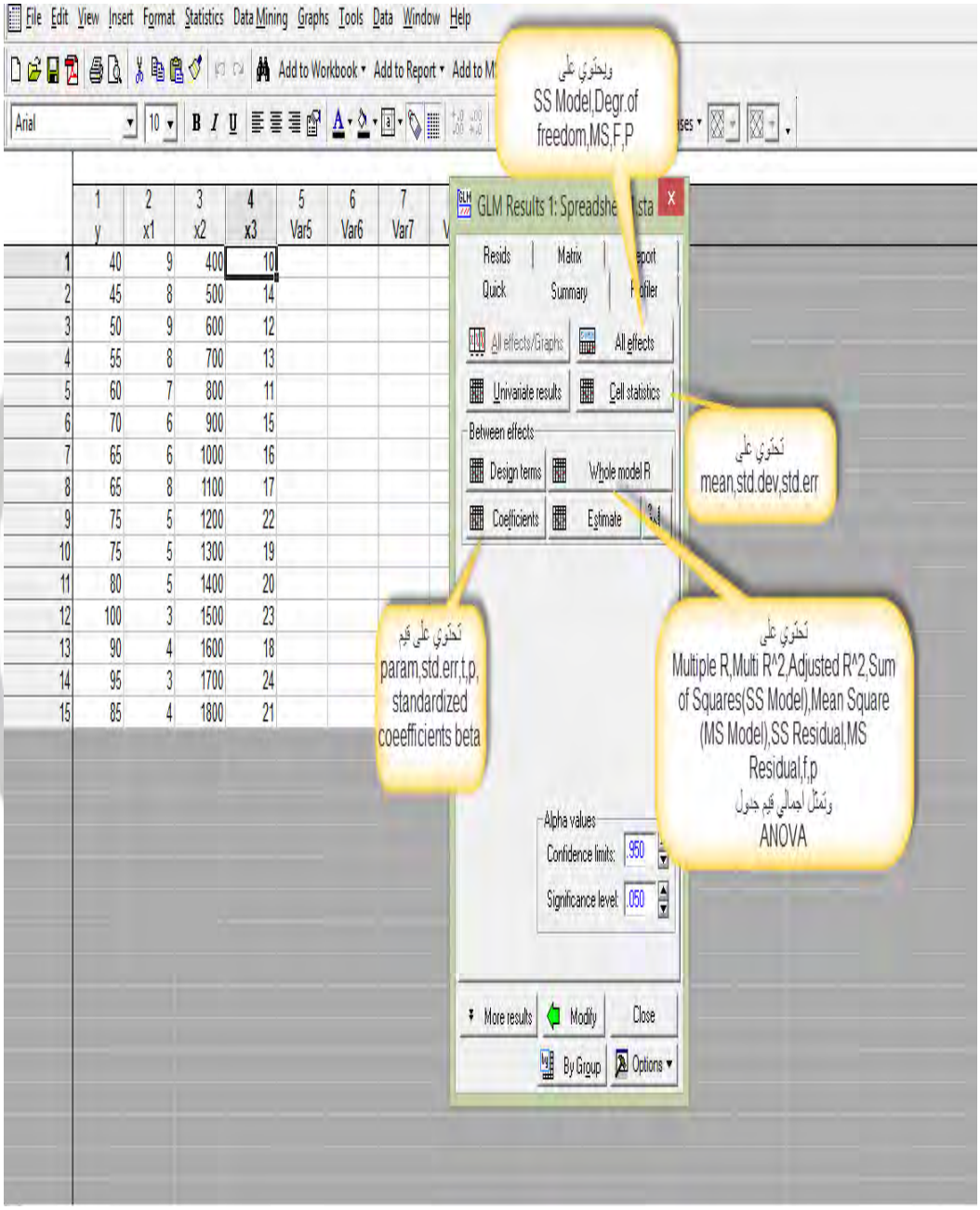


لنحصل على شاشة الحوار ادناه والتي تبين المتغير التابع Y والمتغيرات المستقلة X1,X2,X3 ثم نضغط على OK





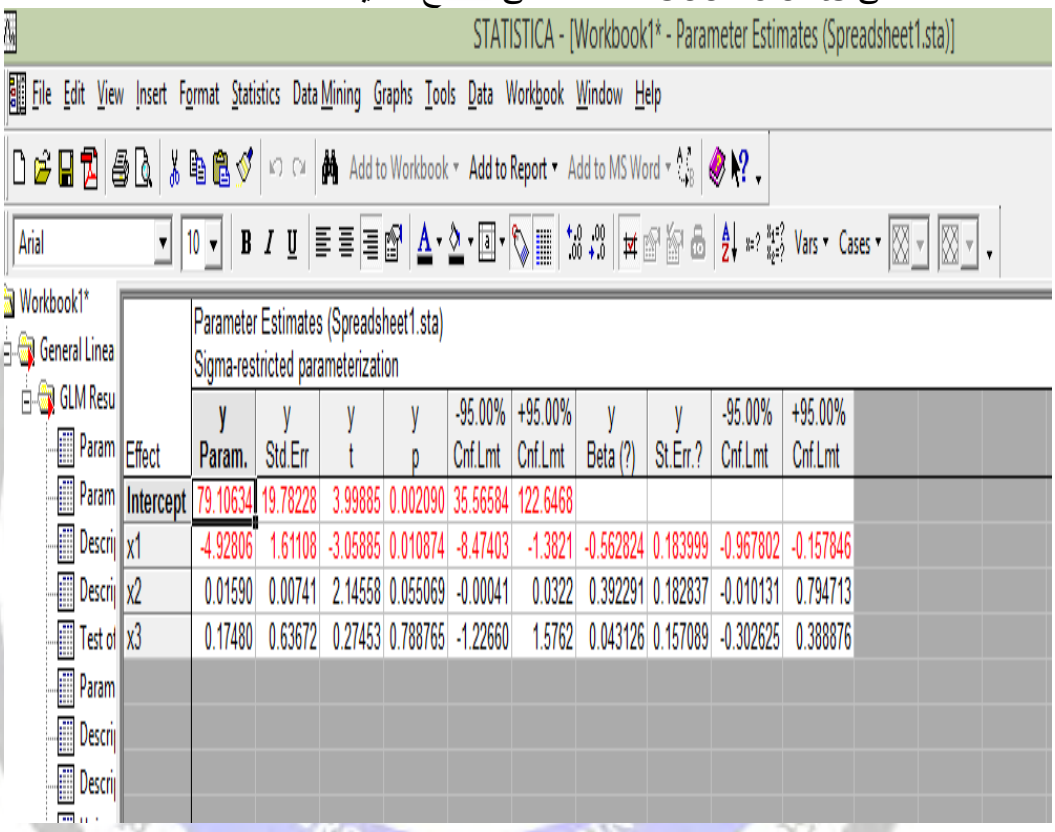
عند الضغط على Summary نحصل على واجهت المخرجات وحسب شاشة الحوار التالية :



	1 y	2 x1	3 x2	4 x3	5 Var5	6 Var6	7 Var7
1	40	9	400	10			
2	45	8	500	14			
3	50	9	600	12			
4	55	8	700	13			
5	60	7	800	11			
6	70	6	900	15			
7	65	6	1000	16			
8	65	8	1100	17			
9	75	5	1200	22			
10	75	5	1300	19			
11	80	5	1400	20			
12	100	3	1500	23			
13	90	4	1600	18			
14	95	3	1700	24			
15	85	4	1800	21			

وعلية تكون النتائج المستخرجة كالاتي :

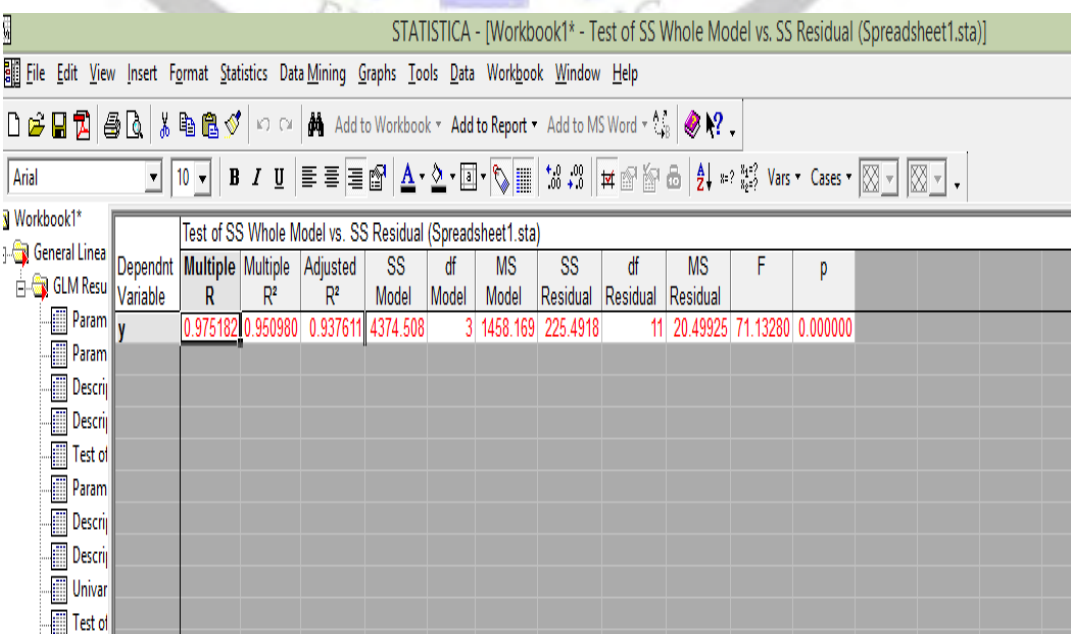
عند الضغط على Coefficients نحصل على النتائج التالية :



Parameter Estimates (Spreadsheet1.sta)
Sigma-restricted parameterization

Effect	y Param.	y Std.Err	y t	y p	-95.00% Cnf.Lmt	+95.00% Cnf.Lmt	y Beta (?)	y St.Err.?	-95.00% Cnf.Lmt	+95.00% Cnf.Lmt
Intercept	79.10634	19.78228	3.99885	0.002090	35.56584	122.6468				
x1	-4.92806	1.61108	-3.05885	0.010874	-8.47403	-1.3821	-0.562824	0.183999	-0.967802	-0.157846
x2	0.01590	0.00741	2.14558	0.055069	-0.00041	0.0322	0.392291	0.182837	-0.010131	0.794713
x3	0.17480	0.63672	0.27453	0.788765	-1.22660	1.5762	0.043126	0.157089	-0.302625	0.388876

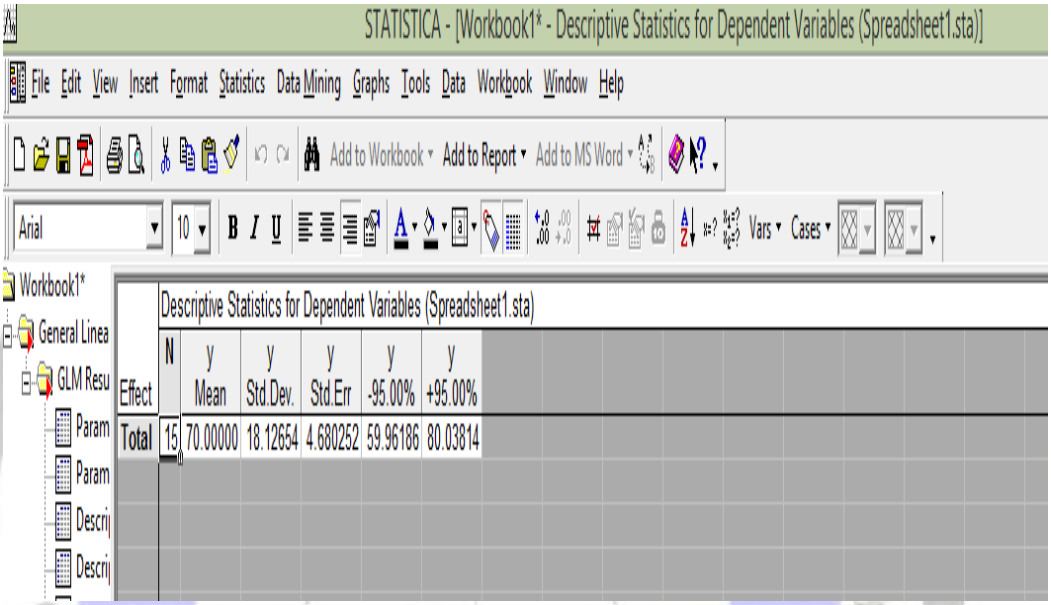
وعند الضغط على Whole model R نحصل على النتائج التالية :



Test of SS Whole Model vs. SS Residual (Spreadsheet1.sta)

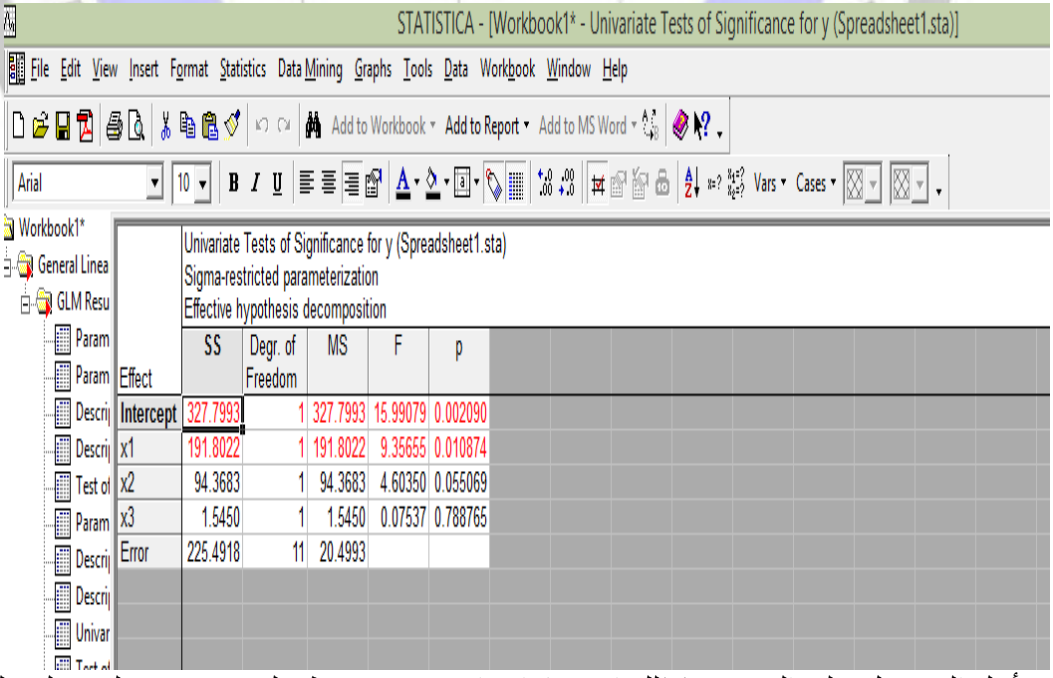
Dependnt Variable	Multiple R	Multiple R^2	Adjusted R^2	SS Model	df Model	MS Model	SS Residual	df Residual	MS Residual	F	p
y	0.975182	0.950980	0.937611	4374.508	3	1458.169	225.4918	11	20.49925	71.13280	0.000000

اما الخيار Cell ststistics فنحصل منه على :



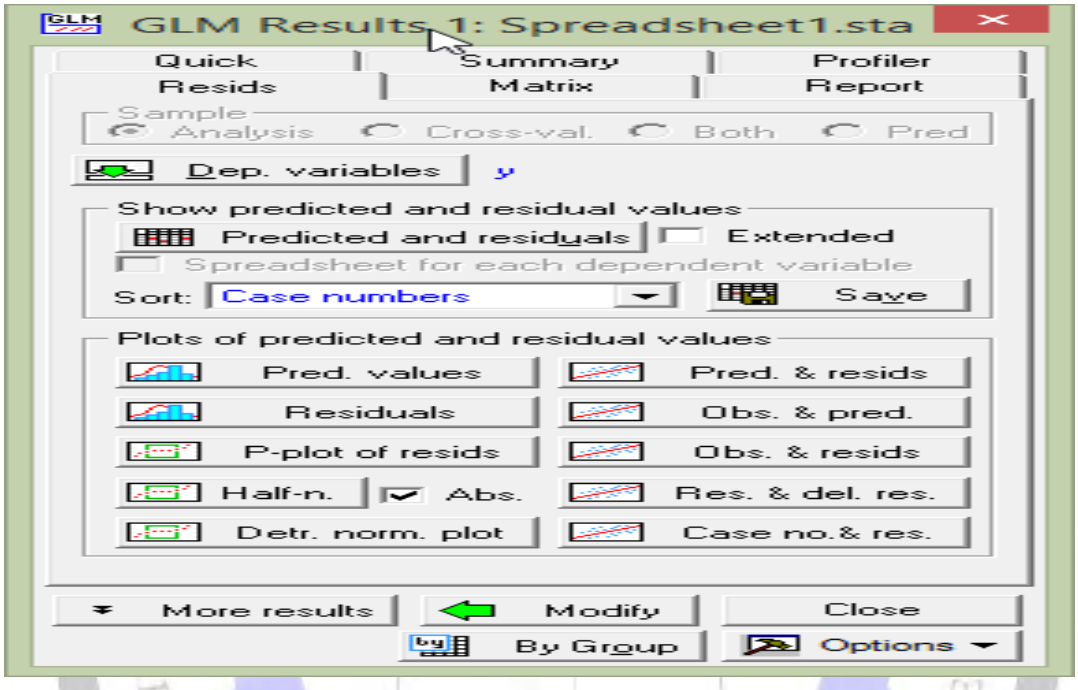
Effect	N	y Mean	y Std.Dev.	y Std.Err	y -95.00%	y +95.00%
Total	15	70.00000	18.12654	4.680252	59.96186	80.03814

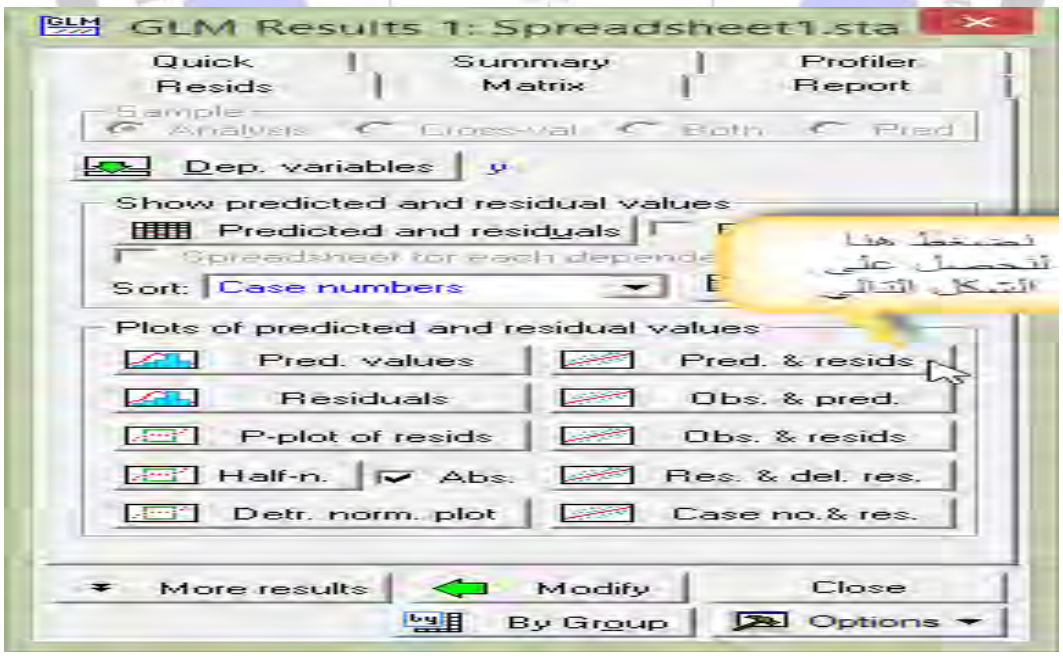
وعند الضغط على All effects نحصل على النتائج التالية :



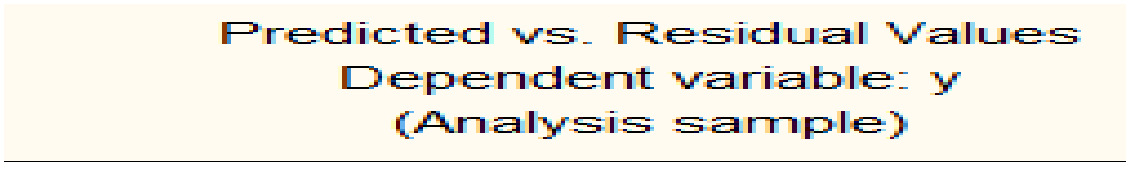
Effect	SS	Degr. of Freedom	MS	F	p
Intercept	327.7993	1	327.7993	15.99079	0.002090
x1	191.8022	1	191.8022	9.35655	0.010874
x2	94.3683	1	94.3683	4.60350	0.055069
x3	1.5450	1	1.5450	0.07537	0.788765
Error	225.4918	11	20.4993		

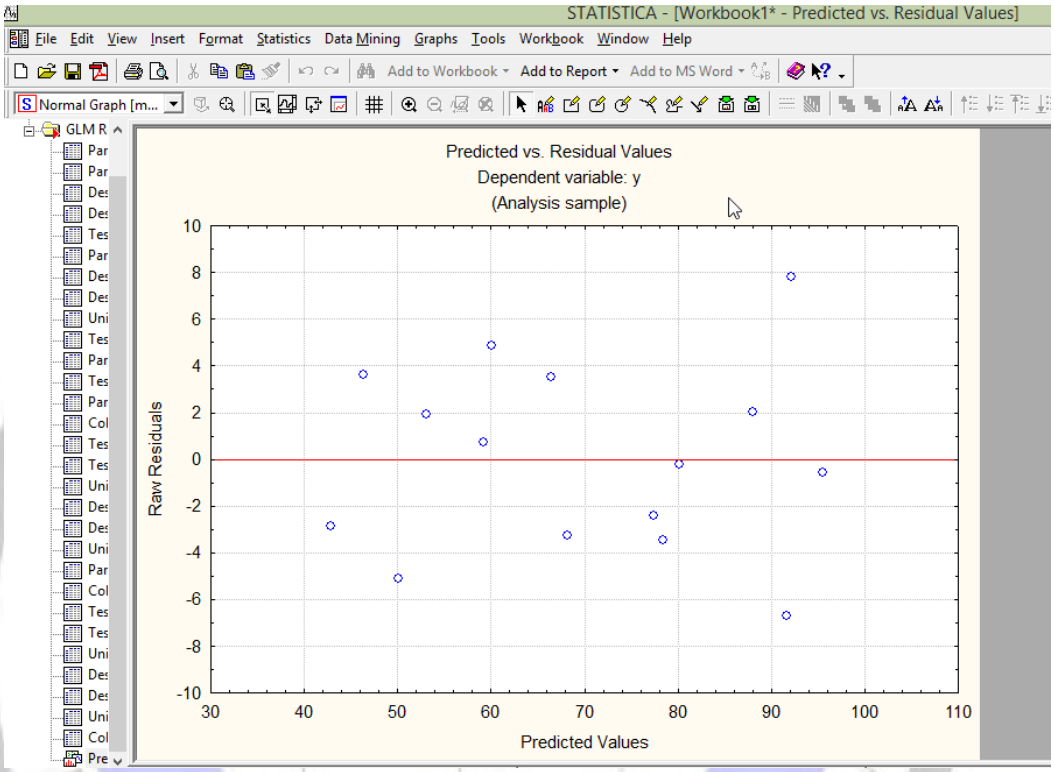
ولأجل الحصول على الرسوم وكذلك الاختبارات الاخرى نضغط على Resids لنحصل على مربع الحوار التالي :





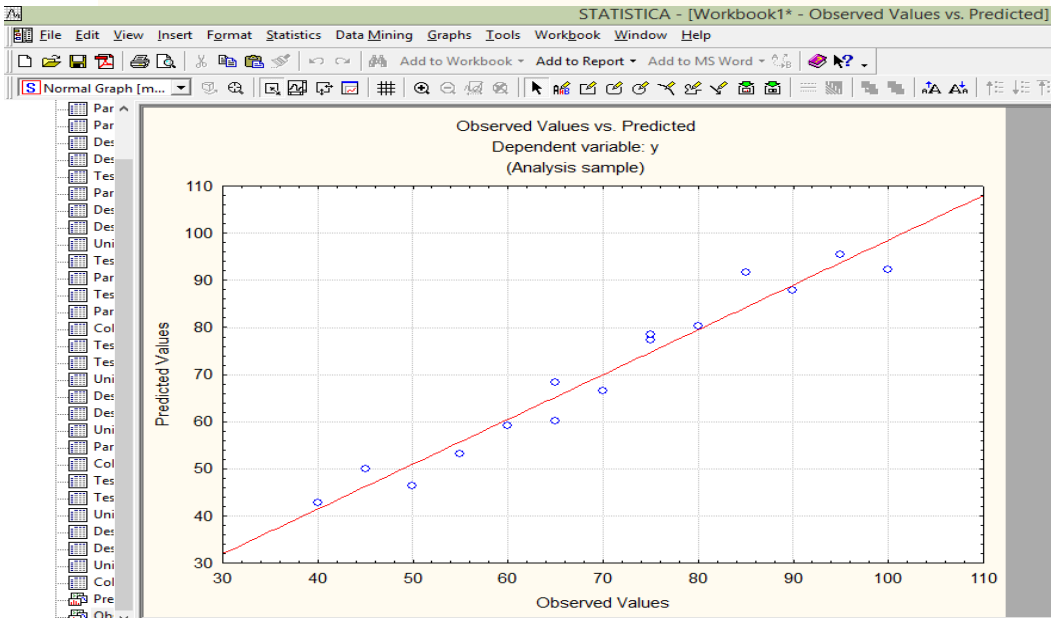
يمثل الشكل علاقة



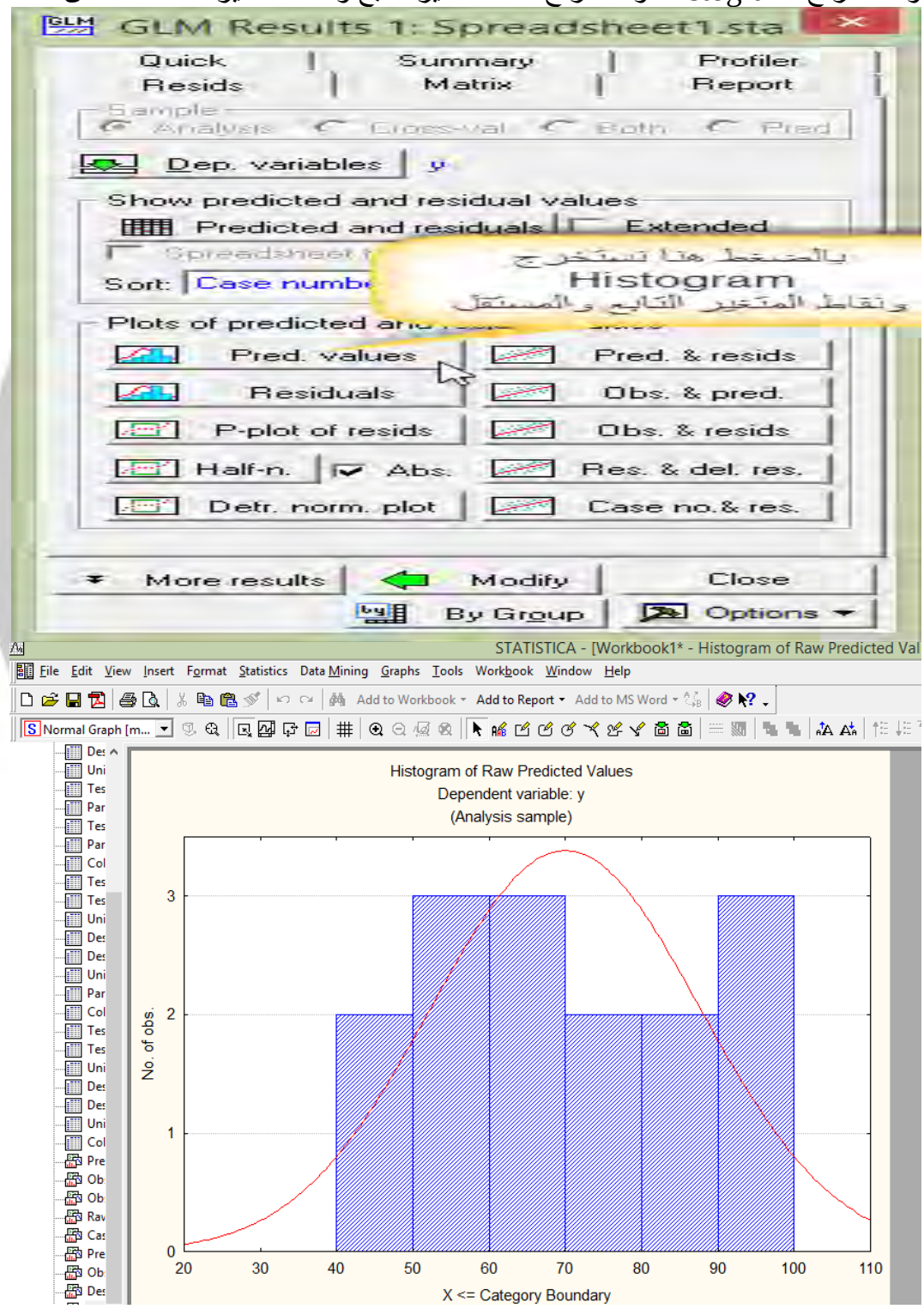


وعند الضغط على زر (obs.&pred) نحصل على الشكل االادناة والذي يوضح العلاقة بين

Observed Values vs. Predicted
Dependent variable: y
(Analysis sample)

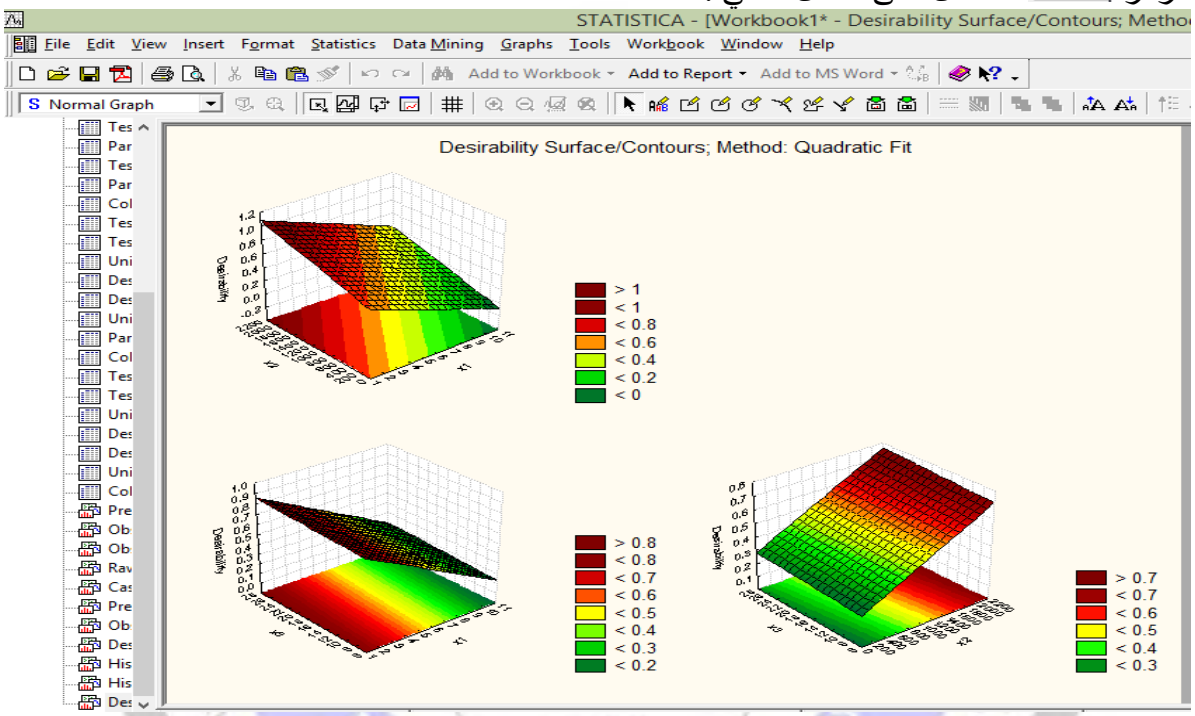


ولاستخراج Histogram واستخراج نقاط المتغير التابع ونقاط المتغيرات المستقلة من خلال



ونطبق الخطوات السابقة التي تم التعرف عليها مسبقا في موضوع رسم Histogram

ولاجل الحصول على الرسومات بصيغة ثلاثية الابعاد نضغط على Profiler ومن ثم نضغط على هذا الرمز لنحصل على الشكل التالي :



تحليل النتائج التي تم الحصول عليها من STATISTICA

نستنتج من المخرجات أعلاه ما يلي :

- من الجدول الأول نلاحظ بأن طريقة الإنحدار المستخدمة وهي طريقة **Enter** حيث يتبين ان البرنامج قام بادخال جميع المتغيرات المستقلة في معادلة الإنحدار الخطي المتعدد .
- من الجدول الثاني نلاحظ بأن قيم معامل الارتباط الثلاثة وهي معامل الارتباط البسيط **R** قد بلغ (0.975) بينما بلغ معامل التحديد R^2 (0.951) في حين كان معامل التحديد المصحح R^{2-} (0.938) مما يعني بأن المتغيرات المستقلة التفسيرية (أختبار الشد لأعلى على العقلة ، أختبار السعة الحيوية ، أختبار الوثب العمودي من الثبات) استطاعت ان تفسر (0.94) من التغيرات الحاصلة في (أختبار الوثب العريض من الثبات) المطلوبة والباقي (0.06) يعزى إلى عوامل اخرى .
- كما يلاحظ في الجدول الثالث بأنه يتضمن قيم تحليل التباين والذي يمكن المعرفة من خلاله على القوة التفسيرية للنموذج ككل عن طريق إحصائية **F** وكما يلاحظ من جدول تحليل التباين

المعنوية العالية لأختبار **F** (**P** < 0.0001) مما يؤكد القوة التفسيرية العالية لنموذج الإنحدار الخطي المتعدد من الناحية الإحصائية .

- أما في الجدول الرابع والأخير فيلاحظ قيمة الثابت ومعاملات الإنحدار ودلالتها الإحصائية للمتغيرات المستقلة على المتغير التابع ويمكن تلخيص هذه الجدول بالشكل الآتي :

المتغير التابع	المتغيرات المستقلة			
Y	***B* الحد الثابت غير المعياري**	**X1**	**X2**	**X3**
قيمة معامل	79.106	- 4.928	0.015	0.175
قيم اختبار T	93.99	-3.059	2.146	0.275
المعنوية	0.002	0.01	0.055	0.789

من الجدول نستنتج ان المتغيرات المستقلة (أختبار الشد لأعلى على العقلة) كان معنوياً من الناحية الإحصائية وحسب أختبار **t** (عند مستوى معنوية **P** ≤ 0.05) ، في حين كاد (أختبار السعة الحيوية) أن يكون معنوياً (عند مستوى معنوية **P** ≤ 0.05) إلا أن المتغير المستقل (أختبار الوثب العمودي من الثبات) لم يكن ذو تأثيراً معنوياً في نموذج الإنحدار المتعدد وحسب أختبار **t** ، ومن الجدول الرابع والأخير يمكن التوصل إلى معادلات الإنحدار باستخدام ***Beta*** غير المعيارية (الحد الثابت) وكما يلي :

- إن معادلة خط إنحدار (الوثب العريض من الثبات) على أختبارات (الشد لأعلى على العقلة والسعة الحيوية والوثب العمودي من الثبات) هي :

الوثب العريض من الثبات = 79.106 + 4.928 × الشد لأعلى على العقلة + 0.015 × السعة الحيوية + 0.175 × الوثب العمودي من الثبات .

- تعد أوزان ***Beta*** المعيارية (الحد الثابت) هي معاملات المسار ***Path coefficients*** (تحليل المسار) حيث يمكن إيجاز تلك المعاملات فيما يلي :

- إدخال (أختبار الوثب العريض من الثبات) كمتغير تابع وأختبارات (الشد لأعلى على العقلة والسعة الحيوية والوثب العمودي من الثبات) كمتغيرات مستقلة

الوثب العريض من الثبات = 0.563 ×الشد لأعلى على العقلة + 0.392 × السعة الحيوية + 0.043 × الوثب العمودي من الثبات .

وهذا السياق خطأ شاع في البحوث والدراسات النفسية والتربوية والإجتماعية إذ يتم استخدام***Beta*** المعيارية لإيجاد معادلات الإنحدار المفترضة .. لذا يجب أن يتم استخدام ***Beta*** غير المعيارية (الحد الثابت).