الانحـــــدار المتـــعدد
Multiple Linear regression

د. سبأ محمد علوان

أستاذ مساعد قسم الاحصاء وبحوث العمليات

جامعة الملك سعود

مقدمة

على الرغم من أهمية استخدام الإحصاء في تحليل البيانات للبحوث والرسائل العلمية فإن هناك بعض المشكلات التي قد تنشأ عن هذا الاستخدام والتي تكون غالباً بدون قصد وبسبب عدم التخصص في الإحصاء لدى البعض وبالتالي عدم الإلمام ببعض الجوانب العلمية الإحصائية الدقيقة .

وفي هذه الورشة نحاول معرفة التحليل الاحصائي المناسب من خلال طريقة سهلة وهي الأمثلة المباشرة والتي تحاكي بعض الحالات التي قد يكون فيها الباحث في مساره البحثي، مع ملاحظة اننا نركز على البحوث العلمية وبما هو متاح من الوقت لهذه الورشة .

مثال :

للمثال رقم 10 جد معاملة الانحدار بطريقة Stepwise Regression

الانحدار التدريجي
Stepwise Regression

للاطلاع

الخطوات النظرية التى تقوم عليها طريقة الانحدار التدريجى

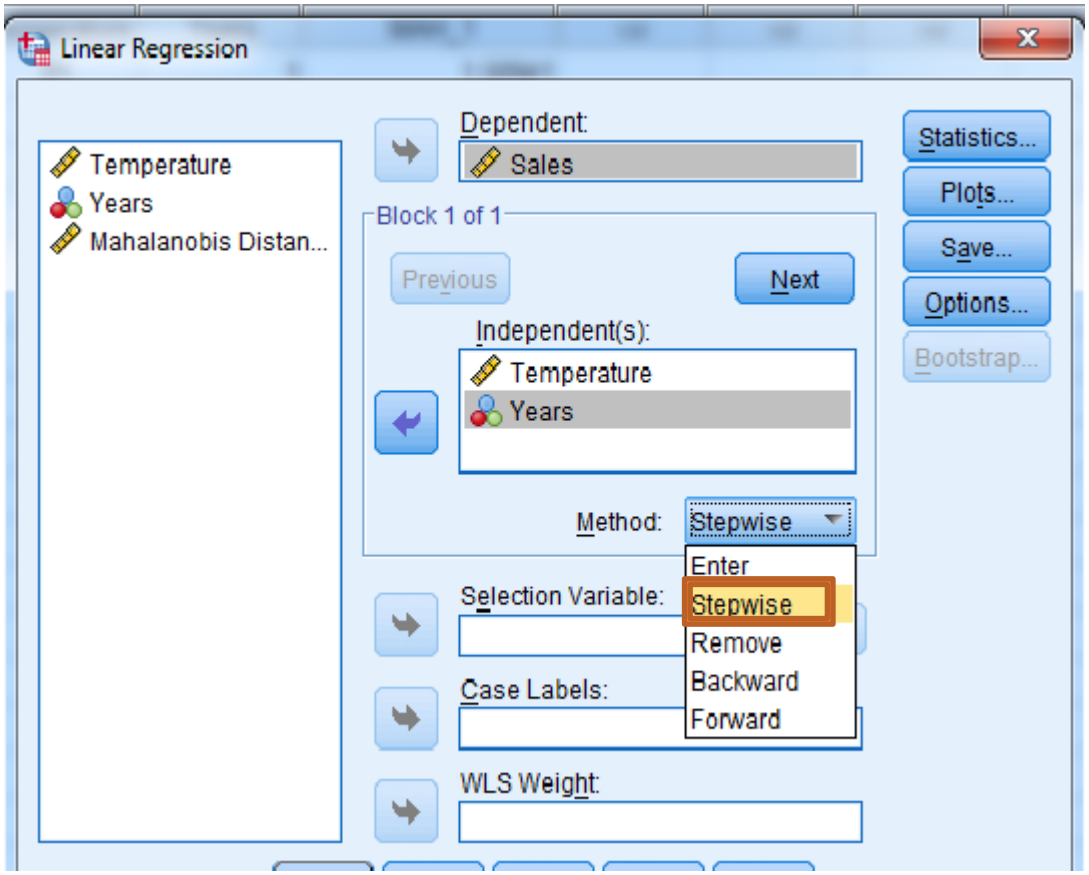
- يبدأ الانحدار التدريجى بالخطوات التالية:

1. نحسب مصفوفة الانحدار لجميع المتغيرات
2. نختار المتغير المستقل الذى له اكبر ارتباط بالمتغير التابع وندخله فى معادله الانحدار
3. نختار المتغير المستقل الثانى الذى له اكبر ارتباط بعد المتغير الذى دخل المعادله فنبحث اولا هل هناك ارتباط كبير بينه وبين المتغير الذى اختير اذا كانت الاجابه بنعم يستبعد ذلك المتغير واذا كانت بلا ندخله فى الاختيار
4. نكرر هذه العمليه مع بقيه المتغيرات المستقله مع استبعاد المتغيرات التى لها ارتباط كبير مع المتغيرات المختاره
5. تكون عملية الاضافه مجديه اذا كان هناك تأثير على معامل التحديد Coefficient of Determination (مربع معامل الارتباط) وكذلك قيمه F من جدول تحليل التباين
6. نتوقف عن الاضافه اذا لم يكن للاضافه تأثير على معامل التحديد وقيمة F المحسوبه (أو له تأثير ضعيف).

خطوات الانحدار التدريجي باستخدام SPSS

خطوات الانحدار التدريجي باستخدام SPSS

هي نفس خطوات الانحدار العياري باستثناء الخطوة التي فيها يتم اختيار الطريقة حيث سنختار Stepwise



النتائج

Variables Entered/Removed[a]			
Model	Variables Entered	Variables Removed	Method
1	Temperature	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

الجدول الاول:

بعنوان Variables Entered/removed ويوضح المتغيرات الداخله فى المعادله وطريقة المعالجه ويتضح أن المتغير temp هو المتغير الوحيد الذى تم ادخاله فى معادله الانحدار.

لاحظ أن Temperature هو المتغير المعنوي في طريقة الانحدار العياري

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.907[a]	.822	.809	6.339

a. Predictors: (Constant), Temperature

b. Dependent Variable: Sales

الجدول الثاني:

بعنوان Model Summary ويعطى بعض المقاييس الهامه والمحسوبه من البيانات وأهمها معامل التوافق (مربع معامل الارتباط) ويستخدم للحكم على عمليه التوفيق ومنه نجد أن معادله الانحدار تمثل 80% من البيانات.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2420.480	1	2420.480	60.228	.000[b]
	Residual	522.453	13	40.189		
	Total	2942.933	14			

a. Dependent Variable: Sales

b. Predictors: (Constant), Temperature

الجدول الثالث:

بعنوان ANOVA وهو تحليل التباين للانحدار ويتضح أن الانحدار معنوى حيث Sig. = 0.00 وهى أقل من 0.05

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-11.271-	6.465		-1.743-	.105
	Temperature	1.600	.206	.907	7.761	.000

a. Dependent Variable: Sales

الجدول الرابع:

بعنوان Coefficients ومنه يمكن ايجاد معادله الانحدار بين المبيعات ودرجة الحرارة فقط ومقدار الخطأ فى التقدير واختبار معنويه المعاملات والتقدير بفتره للمعاملات.

$$Sales = -11.271 + 1.600\,temp$$

Excluded Variables[a]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics: Tolerance
1	Years	.208[b]	.807	.435	.227	.212

a. Dependent Variable: Sales

b. Predictors in the Model: (Constant), Temperature

الجدول الخامس:

بعنوان Excluded Variables ويعرض بيانات تخص المتغير الذى استبعد وهو سنوات الخبره

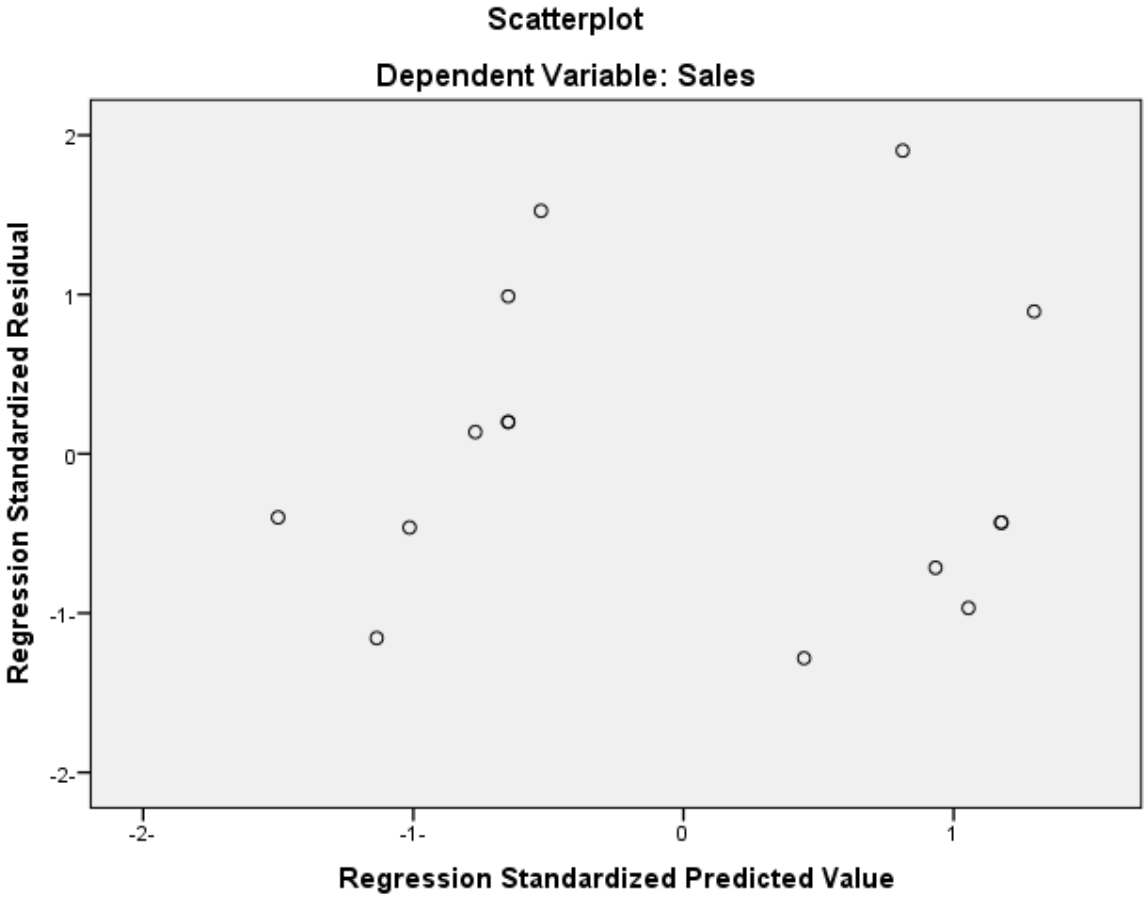
Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	17.53	54.33	37.27	13.149	15
Std. Predicted Value	-1.501-	1.298	.000	1.000	15
Standard Error of Predicted Value	1.803	3.024	2.289	.356	15
Adjusted Predicted Value	18.28	53.28	37.35	13.062	15
Residual	-8.134-	12.066	.000	6.109	15
Std. Residual	-1.283-	1.903	.000	.964	15
Stud. Residual	-1.338-	2.022	-.006-	1.027	15
Deleted Residual	-8.850-	13.613	-.088-	6.947	15
Stud. Deleted Residual	-1.385-	2.346	.021	1.092	15
Mahal. Distance	.199	2.253	.933	.587	15
Cook's Distance	.001	.262	.068	.072	15
Centered Leverage Value	.014	.161	.067	.042	15

a. Dependent Variable: Sales

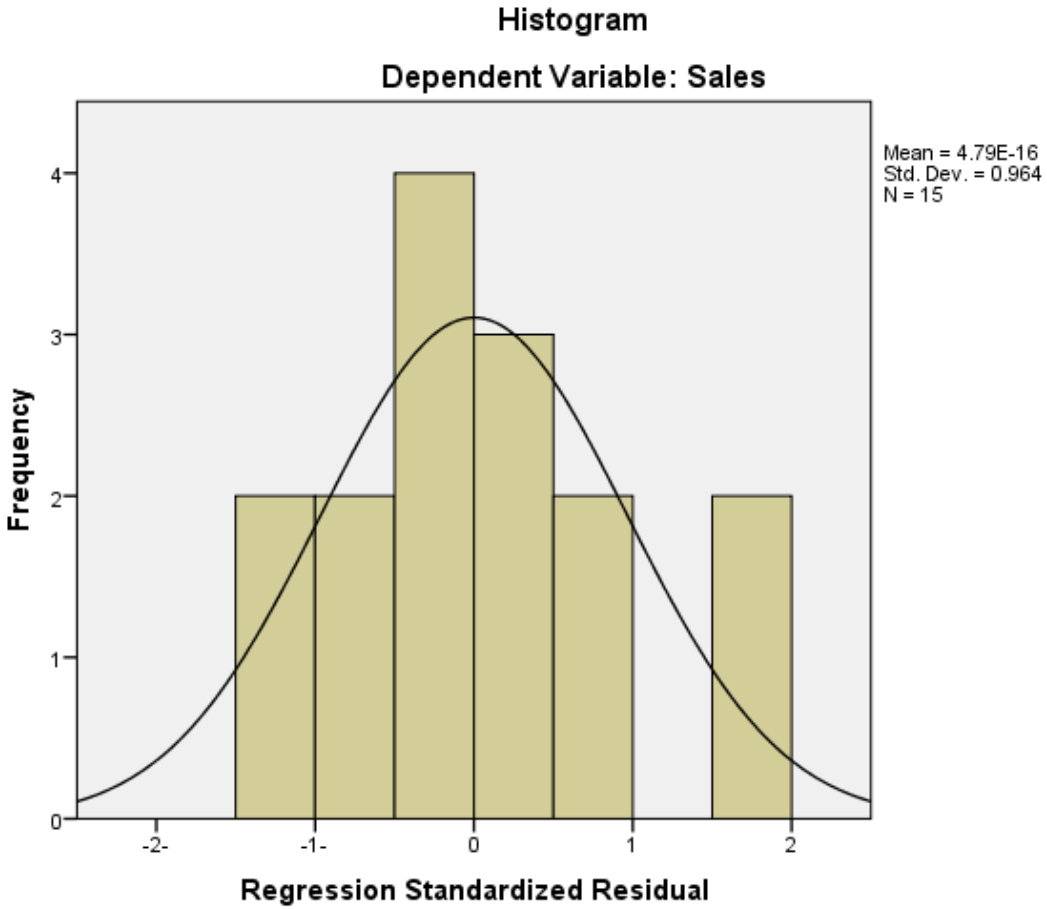
الجدول السادس:

بعنوان Residuals Statistics ويعرض بيانات خاصه بتحليل البواقى.



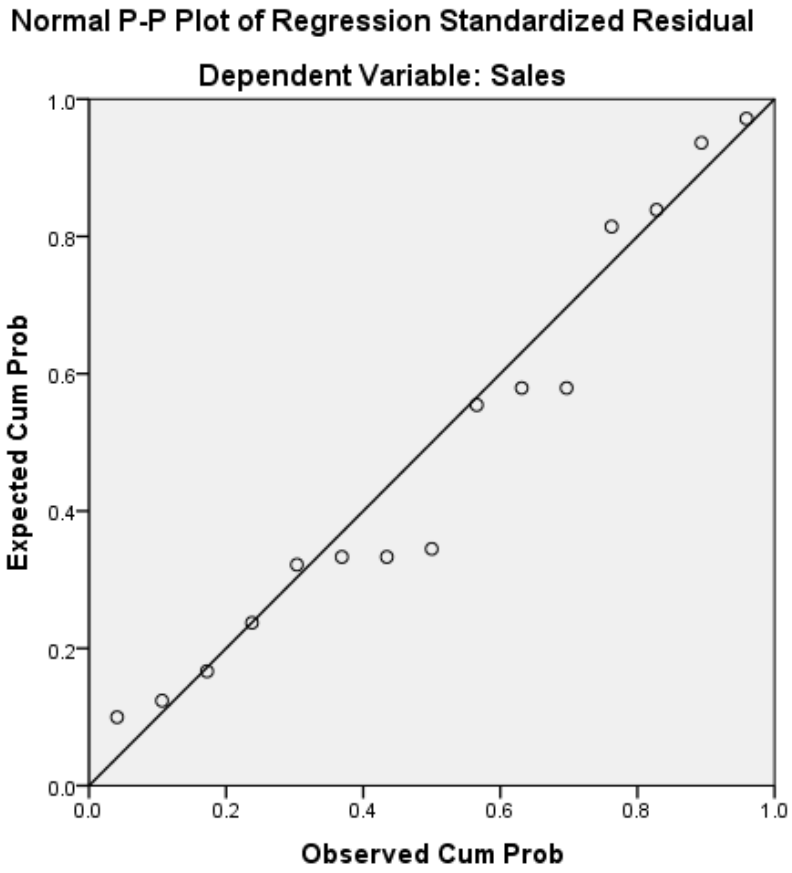
الشكل البيانى الاول:

يمثل شكل الانتشار للبواقى مع القيم المتوقعه ومنه يتضح عدم وجود نمط معين للنقاط فى الشكل وهذا يتسق مع شرط الخطيه.



الشكل البيانى الثالث:

هو المدرج التكرارى ويستخدم للتعرف هل البيانات تتوزع حسب التوزيع الطبيعى أم لا؟



الشكل البيانى الثانى:

يختبر هل البواقى تتبع التوزيع الطبيعى ام لا؟

ومن الشكل نجد أن النقاط تتجمع حول الخط وبالتالى فان البيانات (البواقى) تتوزع حسب التوزيع الطبيعى.

مثال : 12

للمثال رقم 10 جد معاملة الانحدار بطريقة Hierarchical Regression

الانحدار الهرمي
Hierarchical Regression

الانحدار الهرمي
Hierarchical Regression

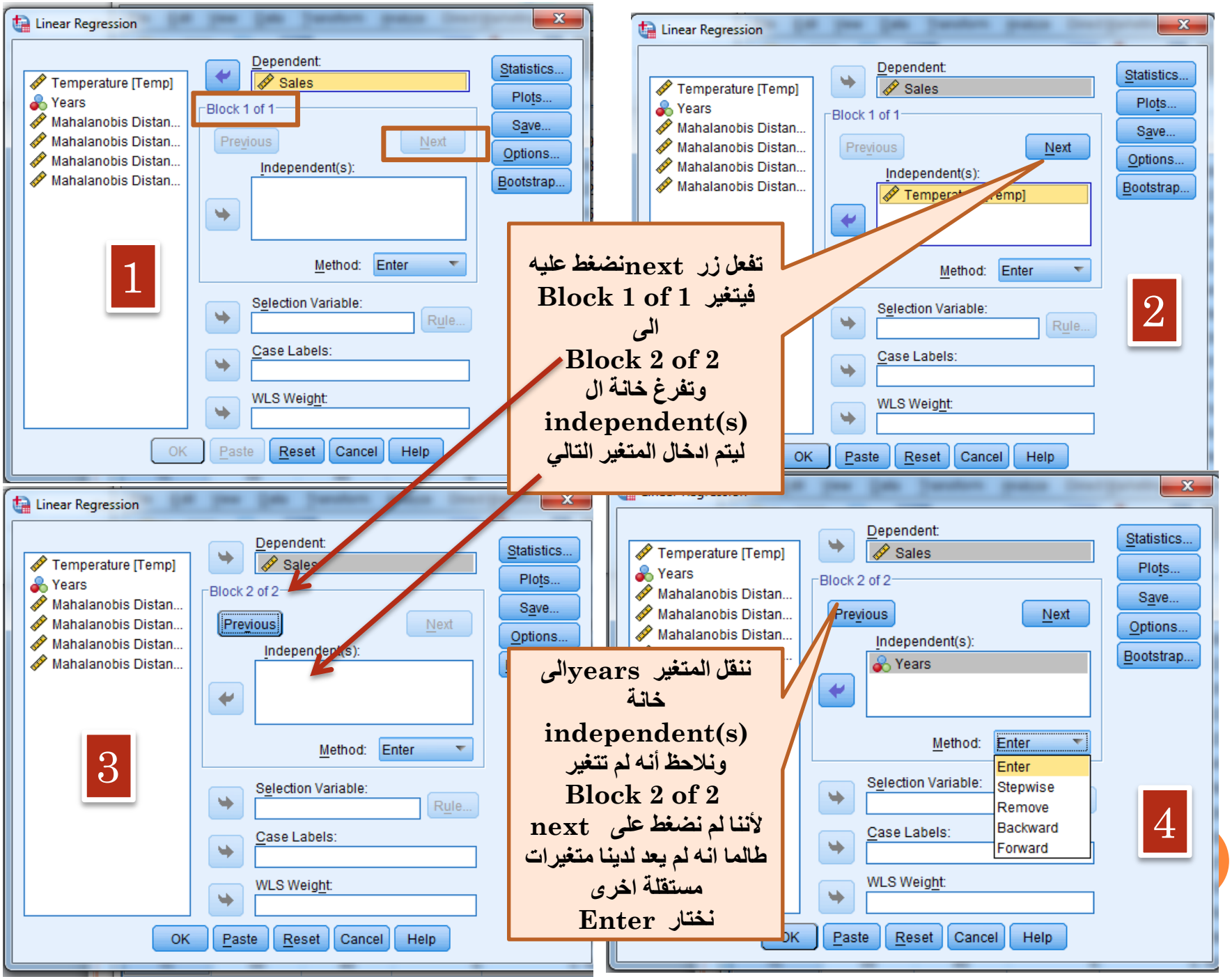
يتم فى هذه الحاله استخدام الطريقة العياريه فى تحديد معادلة الانحدار ولكن ليس لكل المتغيرات بل ندخل المتغيرات تباعا فندخل أول متغير ثم يليه المتغير التالى وهكذا.

يتم التخلص من كل الاختيارات السابقة وابسط الطرق هي حفظ البيانات ومن ثم قفل البرنامج وفتحه من جديد

خطوات الانحدار الهرمي:

Analyze----- Regression------ Linear

لتظهر النافذة التالية


Linear Regression
Temperature [Temp]
Years
Mahalanobis Distan...
Dependent:
Sales
Block 1 of 1
Previous
Next
Independent(s):
Method: Enter
Selection Variable:
Rule...
Case Labels:
WLS Weight:
Statistics...
Plots...
Save...
Options...
Bootstrap...
OK
Paste
Reset
Cancel
Help
1
2
3
4
تفعل زر next نضغط عليه
فيتغير Block 1 of 1
الى
Block 2 of 2
وتفرغ خانة ال
independent(s)
ليتم ادخال المتغير التالي
ننقل المتغير years الى
خانة
independent(s)
ونلاحظ أنه لم تتغير
Block 2 of 2
لأننا لم نضغط على next
طالما انه لم يعد لدينا متغيرات
مستقلة اخرى
نختار Enter
Enter
Stepwise
Remove
Backward
Forward

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Temperature[b]	.	Enter
2	Years[b]	.	Enter

a. Dependent Variable: Sales

b. All requested variables entered.

الجدول الاول:

بعنوان Variables Entered/Removed يوضح المتغيرات التى ادخلت للنموذج وطريقة الاختيار

Model Summary[c]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.907[a]	.822	.809	6.339	.822	60.228	1	13	.000
2	.912[b]	.832	.804	6.426	.009	.652	1	12	.435

a. Predictors: (Constant), Temperature

b. Predictors: (Constant), Temperature, Years

c. [

الجدول الثانى:

بعنوان Model Summary ويعطى ملخص عن النموذج لاحظ كلا من R square Change, F, Sig. نجد أن اضافة المتغير الأول temp معنوى وله جدوى بخلاف المتغير الثانى years

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2420.480	1	2420.480	60.228	.000[b]
	Residual	522.453	13	40.189		
	Total	2942.933	14			
2	Regression	2447.388	2	1223.694	29.633	.000[c]
	Residual	495.546	12	41.295		
	Total	2942.933	14			

a. Dependent Variable: Sales

b. Predictors: (Constant), Temperature

c. Predictors: (Constant), Temperature, Years

الجدول الثالث:

بعنوان ANOVA ويعطى تحليل التباين لكل خطوه ادخال ومنه يتضح معنويه الانحدار فى كل خطوه.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-11.271-	6.465		-1.743-	.105		
	Temperature	1.600	.206	.907	7.761	.000	1.000	1.000
2	(Constant)	-6.993-	8.428		-.830-	.423		
	Temperature	1.275	.454	.722	2.807	.016	.212	4.721
	Years	1.907	2.362	.208	.807	.435	.212	4.721

الجدول الرابع:

بعنوان Coefficients ومنه يحسب خط الانحدار المقترح (المقدر) واختبار معنويه المعاملات وخطأ التقدير. ومن الجدول نجد أن:

1. الخطوة الاولى تم ادخال درجات الحرارة temp وكان الانحدار معنوى.
2. الخطوة الثانية تم ادخال سنوات الخبره مع درجات الحرارة فكانت غير معنويه

وعلى ذلك يكون أفضل خط مقدر هو

$$Sales = -11.271 + 1.600\,temp$$

Excluded Variables[a]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	Years	.208[b]	.807	.435	.227	.212	4.721	.212

a. Dependent Variable: Sales

b. Predictors in the Model: (Constant), Temperature

الجدول الخامس:

بعنوان Excluded Variables وهو خاص بالمتغيرات المستبعده

أما بقية النتائج فهي مشابهة لما سبق

مثال : 13

حيث أن متغير sales لم يكن ارتباطه الخطي ب year معنوي فماهي العلاقة بينهما إذن؟

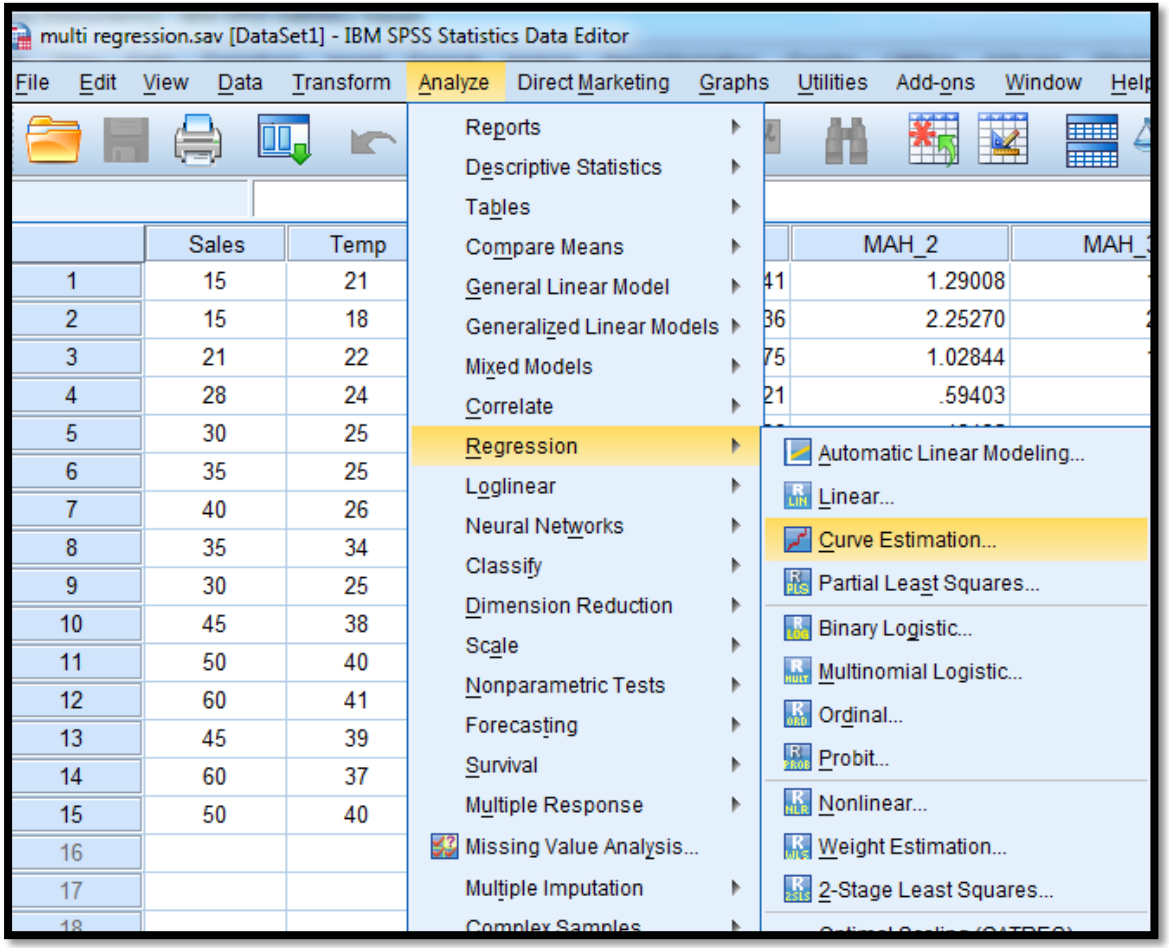
الانحدار غير الخطي
Curve Regression

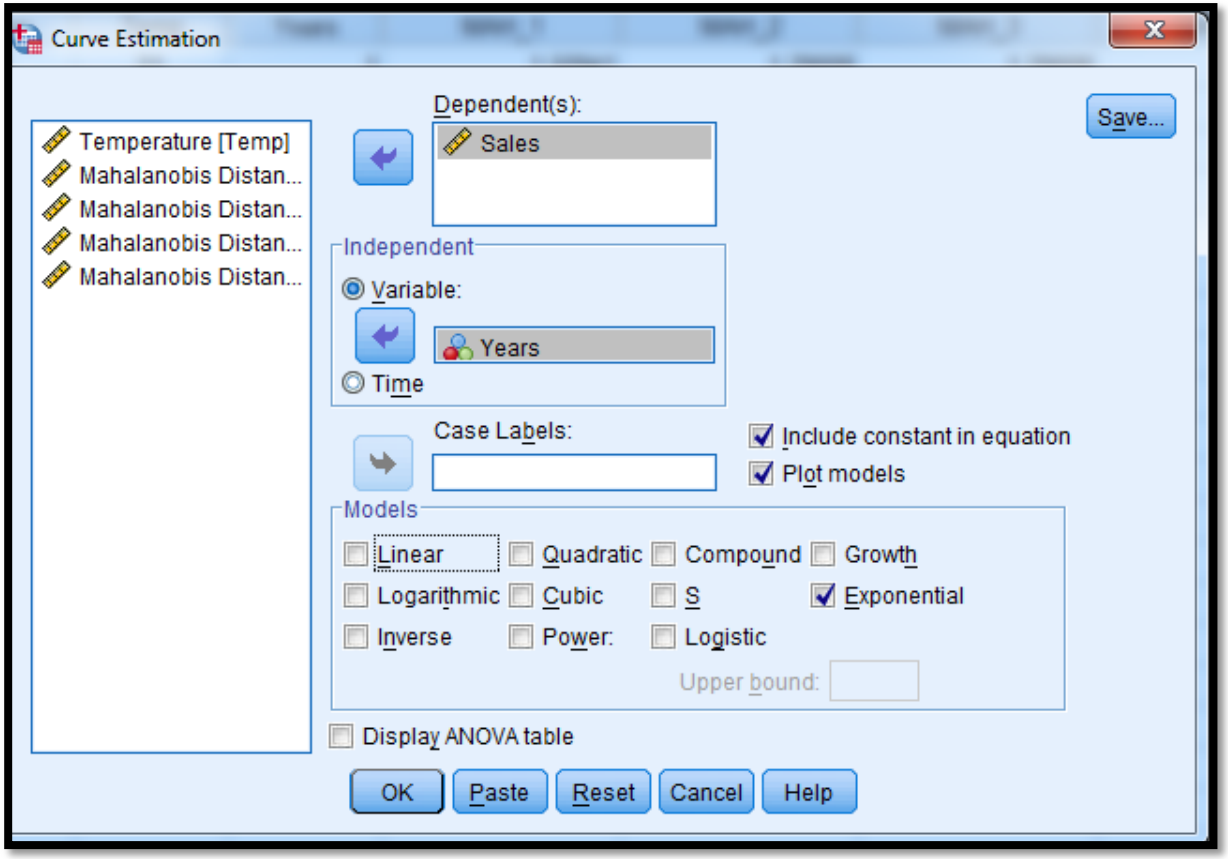
الانحدار غير الخطي
Curve Regression

- بفرض أن لدينا متغيرين وأن العلاقة بينهما غير خطيه. ونقصد بكلمة غير خطيه أن العلاقة تأخذ أى صورة من صورة العلاقات الغير خطيه مثل الدرجة الثانيه أو الثالثة ... الخ.
- وقد تكون العلاقة بينهما أسيه أو لوغاتيميه أو ... الخ.
- وتساعد الحزمة SPSS على ايجاد معادلة الانحدار فى حالة الانحدار الغير خطى وكذلك اجراء الاختبارات المعنويه اللازمة للتقدير وذلك لعدة نماذج غير خطيه منها:

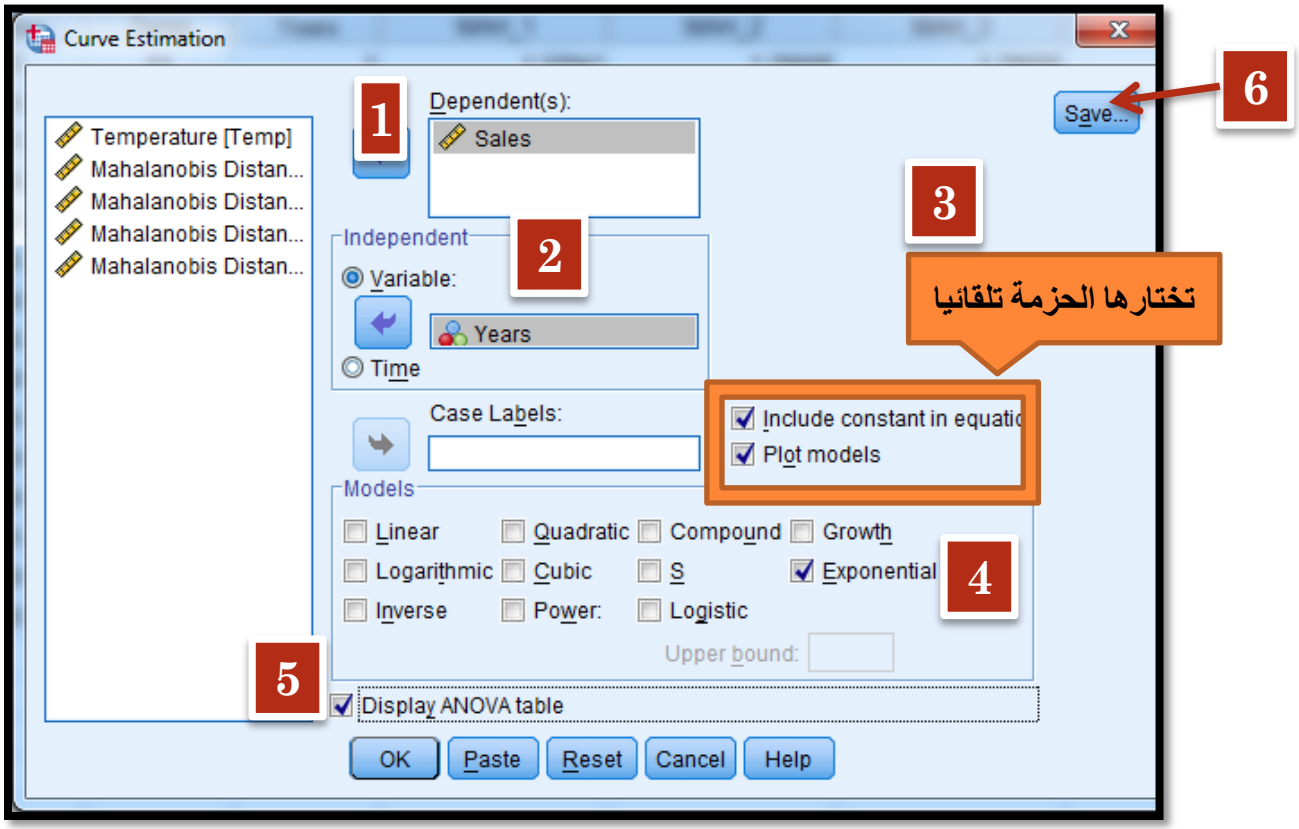
1. العلاقات الخطيه Linear ومن الدرجة الثانيه Quadratic والثالثة Cubic
2. العلاقات الآسيه Exponential
3. العلاقات اللوغاتيميه Logarithmic
4. العلاقات المعكوسه Inverse والعلاقات المركبه Compound
5. العلاقات اللوجستيه Logistic وعلاقة القوة Power

الخطوات:


multi regression.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Automatic Linear Modeling...
Linear...
Curve Estimation...
Partial Least Squares...
Binary Logistic...
Multinomial Logistic...
Ordinal...
Probit...
Nonlinear...
Weight Estimation...
2-Stage Least Squares...
Sales Temp MAH_2
1 15 21 1.29008
2 15 18 2.25270
3 21 22 1.02844
4 28 24 .59403
5 30 25
6 35 25
7 40 26
8 35 34
9 30 25
10 45 38
11 50 40
12 60 41
13 45 39
14 60 37
15 50 40
16
17


Curve Estimation
Temperature [Temp]
Mahalanobis Distan...
Mahalanobis Distan...
Mahalanobis Distan...
Mahalanobis Distan...
Dependent(s):
Sales
Save...
Independent
Variable:
Years
Time
Case Labels:
Include constant in equation
Plot models
Models
Linear
Quadratic
Compound
Growth
Logarithmic
Cubic
S
Exponential
Inverse
Power:
Logistic
Upper bound:
Display ANOVA table
OK
Paste
Reset
Cancel
Help

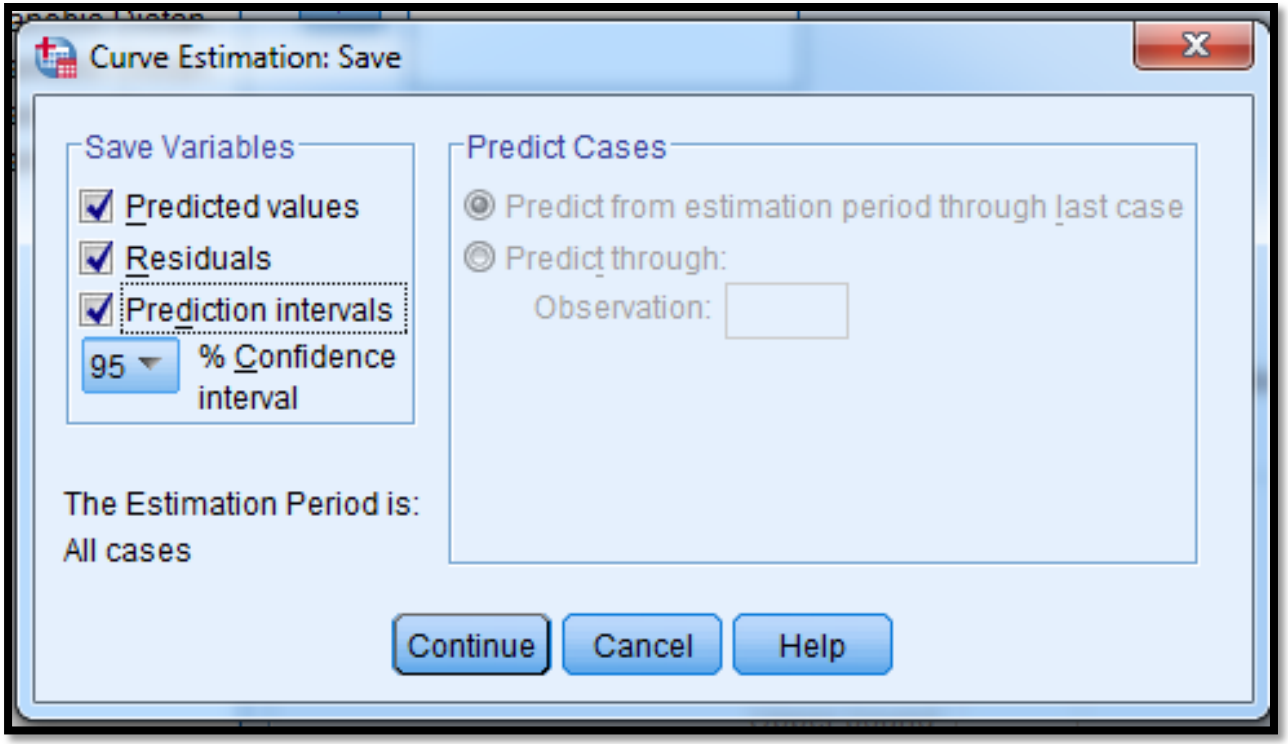
تظهر شاشة جديده بعنوان Curve Estimation ننقل المتغير التابع Sales لخانة Dependent والمتغير المستقل years لخانة Independent مع ملاحظه الاختيار Variable



نختار شكل المنحنى المطلوب وليكن Exponential

نلاحظ ان الامرين Include constant in Equation, Plot Models تم اختيارهما من قبل الحزمه.

تظهر شاشة جديده بعنوان Curve Estimation: Save نختار منها Predicted Values, Residuals, Predicted Interval ونلاحظ حدود الثقة المختارة هى 95% ويمكن تعديلها



نضغط على Continue فنعود للشاشه السابقه

نضغط على Ok فتظهر النتائج التاليه:

Model Description

Model Name		MOD_1
Dependent Variable	1	Sales
Equation	1	Exponential[a]
Independent Variable		Years
Constant		Included
Variable Whose Values Label Observations in Plots		Unspecified

a. The model requires all non-missing values to be positive.

الجدول الأول:

بعنوان Model Description ويحتوى على معلومات خاصه بالنموذج المحدد

اسم المتغير التابع والمستقل ونوع النموذج.

Sales

Exponential

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.832	.692	.669	.256

The independent variable is Years.

الجدول الرابع:

بعنوان Model Summary ويعطى معامل الارتباط ومربع معامل الارتباط

وايضا مربع معامل الارتباط المصحح وخطأ التقدير.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	1.916	1	1.916	29.237	.000
Residual	.852	13	.066		
Total	2.767	14			

The independent variable is Years.

الجدول الخامس:

بعنوان ANOVA ويعطى تحليل التباين لنموذج الانحدار المحدد (الانحدار الآسى) ونجد أن الانحدار معنوى.

Coefficients

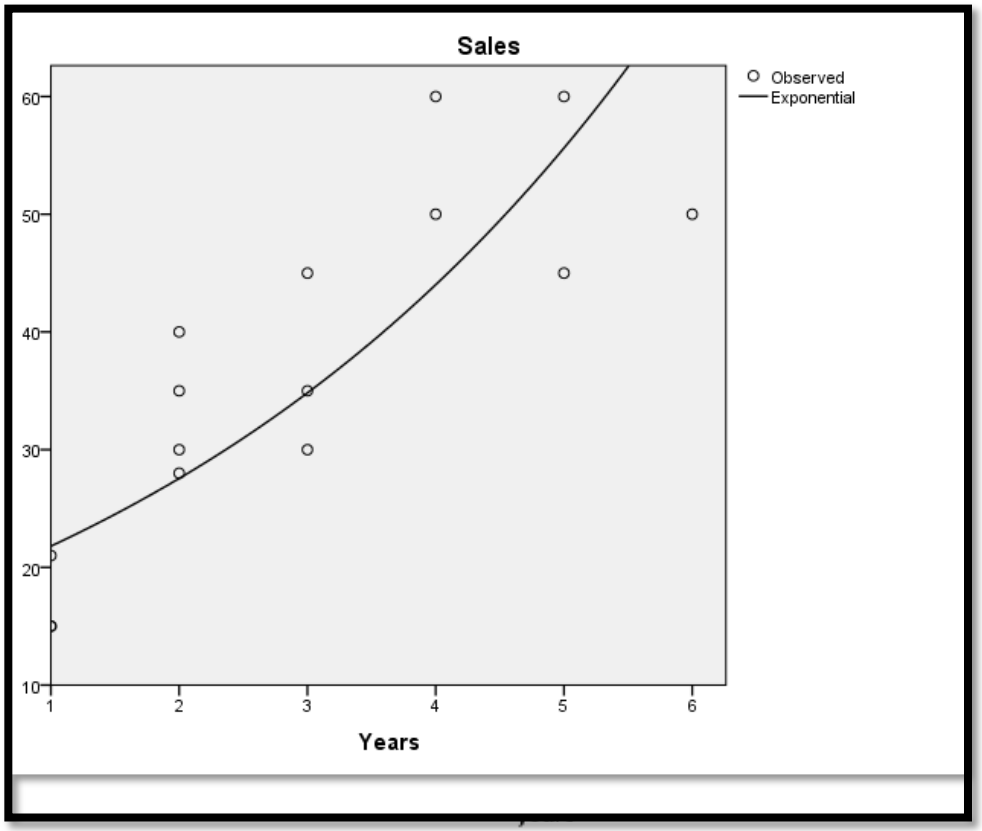
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Years	.234	.043	.832	5.407	.000
(Constant)	17.250	2.470		6.983	.000

The dependent variable is ln(Sales).

الجدول السادس:

بعنوان Coefficients ويعطى قيم معاملات نموذج الانحدار ومعنوية كل معامل على حده.

$$sales = 17.25\, e^{0.23419\, years}$$



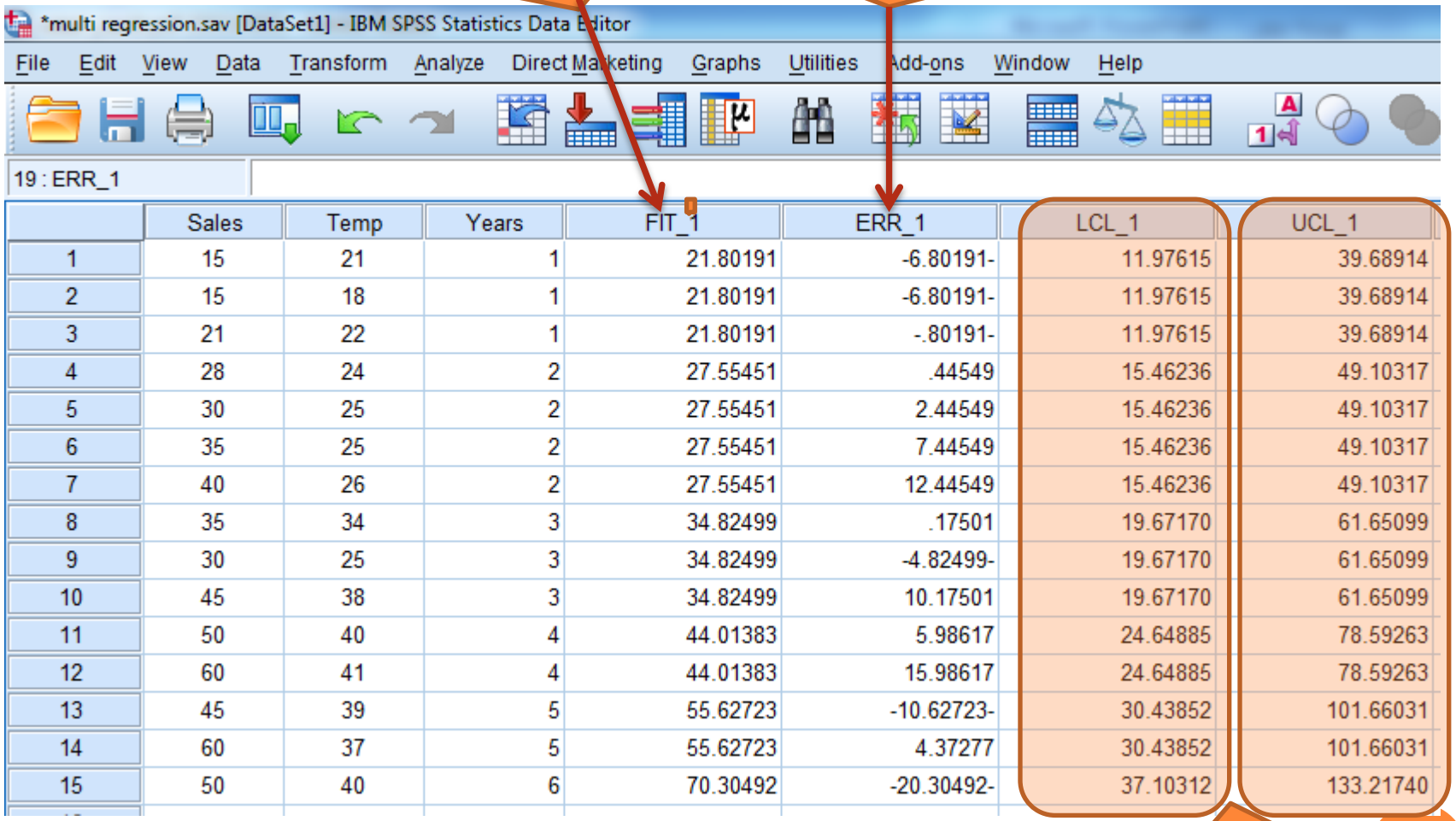
الشكل البياني: شكل الانتشار للبيانات والمنحنى المقدر للبيانات من نموذج الانحدار.

- بالعودة للملف regression.sav نجد أن الحزمة قد أضافت بعض المتغيرات الجديده

بالعودة الى ملف البيانات سنجد اضافة اربعة اعمدة

القيم المقدرة للمتغير Sales من النموذج
مثلا اول قيمة = (1) 17.25e^0.23419

الاخطاء
اول خطأ =15-21.80191



	Sales	Temp	Years	FIT_1	ERR_1	LCL_1	UCL_1
1	15	21	1	21.80191	-6.80191-	11.97615	39.68914
2	15	18	1	21.80191	-6.80191-	11.97615	39.68914
3	21	22	1	21.80191	-.80191-	11.97615	39.68914
4	28	24	2	27.55451	.44549	15.46236	49.10317
5	30	25	2	27.55451	2.44549	15.46236	49.10317
6	35	25	2	27.55451	7.44549	15.46236	49.10317
7	40	26	2	27.55451	12.44549	15.46236	49.10317
8	35	34	3	34.82499	.17501	19.67170	61.65099
9	30	25	3	34.82499	-4.82499-	19.67170	61.65099
10	45	38	3	34.82499	10.17501	19.67170	61.65099
11	50	40	4	44.01383	5.98617	24.64885	78.59263
12	60	41	4	44.01383	15.98617	24.64885	78.59263
13	45	39	5	55.62723	-10.62723-	30.43852	101.66031
14	60	37	5	55.62723	4.37277	30.43852	101.66031
15	50	40	6	70.30492	-20.30492-	37.10312	133.21740

الحدود العليا والدنيا لفترات الثقة للمتغير التابع عند كل قيمة للمستقل

المراجـــــــــــــــع

١)ابو سريع، رضا.(٢٠٠٤).تحليل البيانات باستخدام برنامج spss ،دار الفكر،عمان.

٢)البشير، سعد.(٢٠٠٣)،دليلك الى البرنامج الاحصائي spss ،المعهد العربي للتدريب والبحوث الاحصائية ،العراق

٣)الاختبارات الاحصائية البارامترية واللابارامترية باستخدام SPSSالاصدار السابع عشر ، د.سوسن ابراهيم أبو العلا شلبي، جامعة الملك سعود

٤)الارتباط والانحدار د. كامل أبو ضاهر ، الجامعة الاسلامية –غزة

٥)-بعض المعادلات الإحصائية المستخدمة في تحديد عينة البحث، الهزاع، هزاع محمد. فسيولوجيا الجهد البدني: الأسس النظرية والإجراءات المعملية للقياسات الفسيولوجية. الفصل الرابع. كتاب مقدم للنشر.

٦)Multiple Regression, Abdelfatah Mustafa, Mansuora University

http://www.nca.umich.edu/sample_size_chart
http://www.surveysystem.com/sscalc.htm