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 : 10

سنفترض البيانا ت التالية وهي ل ١٥ محل تجاري كعينة عشوائية في يوم معين ولسلعة ما، اليوم ودرجات الحرارة المسجلة في منتصف اليوم وسنوات الخبرة للبائع الذي يقوم بالخدمة .

لتكن:

- المبيعات Sales هي المتغير التابع dependent variable
- سنوات الخبره , Years درجة الحرارة Temperature هما المتغيراتن المستقلان،
- وسنستخدم الرموز التالية:
  - Y=Sales
  - *X*1=Temperature
  - X2= Years

المطلوب بناء نموذج خطي يصف ويتنبأ بحجم المبيعات بمعلومية درجة الحرارة وسنوات الخبرة للبائع

Sales	Temperature	Years
15	21	1
15	18	1
21	22	1
28	24	2
30	25	2
35	25	2
40	26	2
35	34	3
30	25	3
45	38	3
50	40	4
60	41	4
45	39	5
60	37	5
50	40	6

إن النموذج الخطي المقترح له الشكل التالي :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

معالم مجهولة Parameters

الخطأ في النموذج

وعلى خطأ النموذج بعض الشروط وهي

1. مجموع الاخطاء العشوائية مساويا للصفر وبذلك يكون متوسط الأخطاء مسويا الصفر
2. تباين الخطأ مقدار ثابت لكل المشاهدات ومساويا σ^2 هناك تجانس
3. التغاير بين اى خطأين مساويا الصفر وبذك لا يوجد ارتباط بين الاخطاء وبعضها البعض
4. يفترض أن الخطأ يتوزع حسب التوزيع الطبيعى بمتوسط صفر وتباين مشترك σ^2

كيف يمكن الحصول على هذه المعادلة؟

يوجد ثلاثة أنواع من نماذج الانحدار :

الانحدار العياري : Standard or Simultaneous Regression

وفيها يتم ادخال **كل** المتغيرات المستقلة دفعة واحدة .
ولا نتعرض لمناقشة هل المتغيرات المستقلة مرتبطة ببعضها البعض ام مستقلة

الانحدار االهرمي : Hierachical Regression

وفي هذا النوع ندخل المتغيرات تباعا (**واحدا واحدا**)

الانحدار ال: Stepwise Regression

وفي هذا النوع ندخل **عدد** من المتغيرات (**وليس كل** المتغيرات).
وقد يكون الادخال

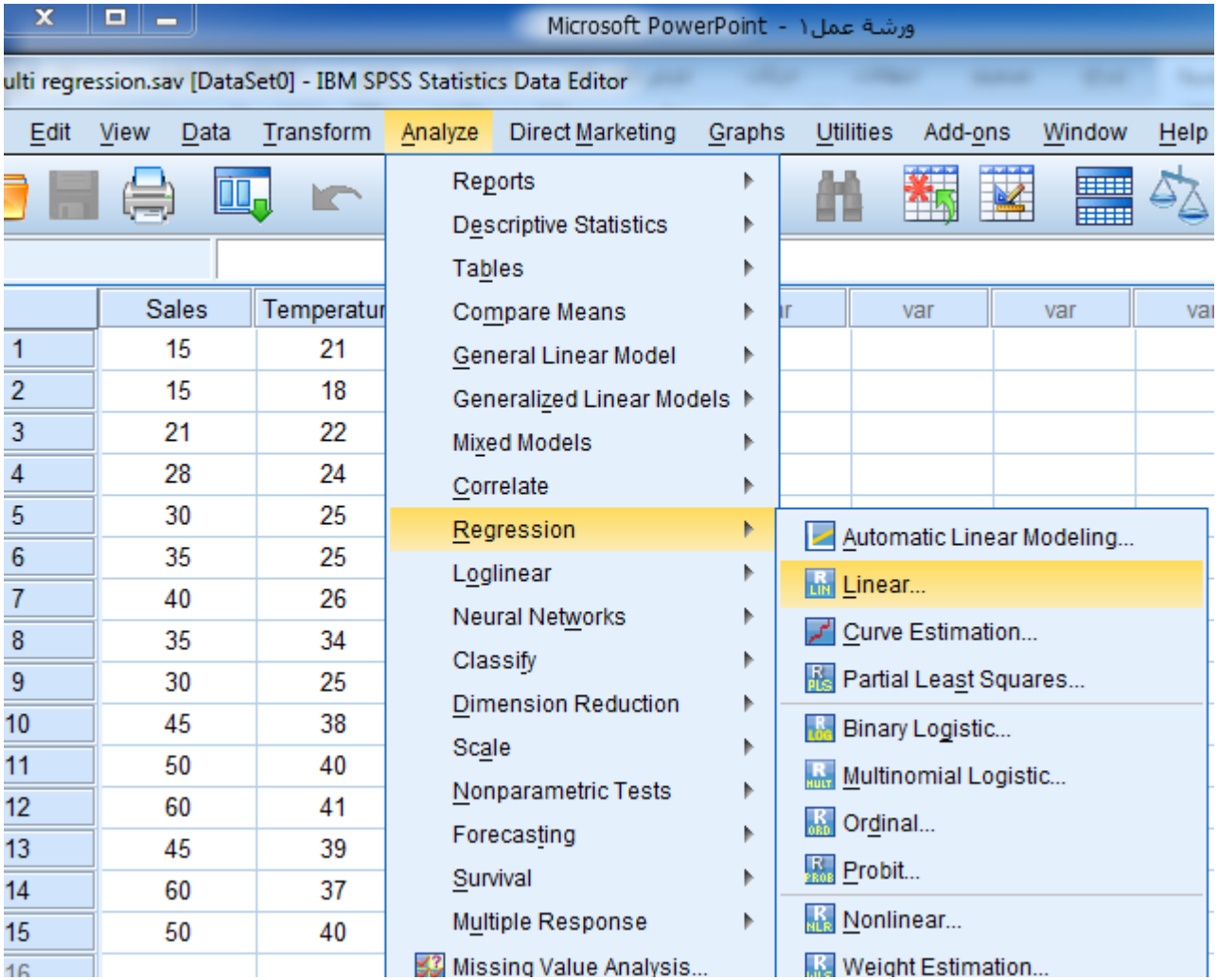
للأمام (forward)
للخلف (backward)
مزج بين الأسلوبين

- ويمكن الان التعرض لبعض انواع الانحدار وكيفيه استخدام الحزمة SPSS فى ذلك وسوف نهتم هنا بالانواع التاليه:

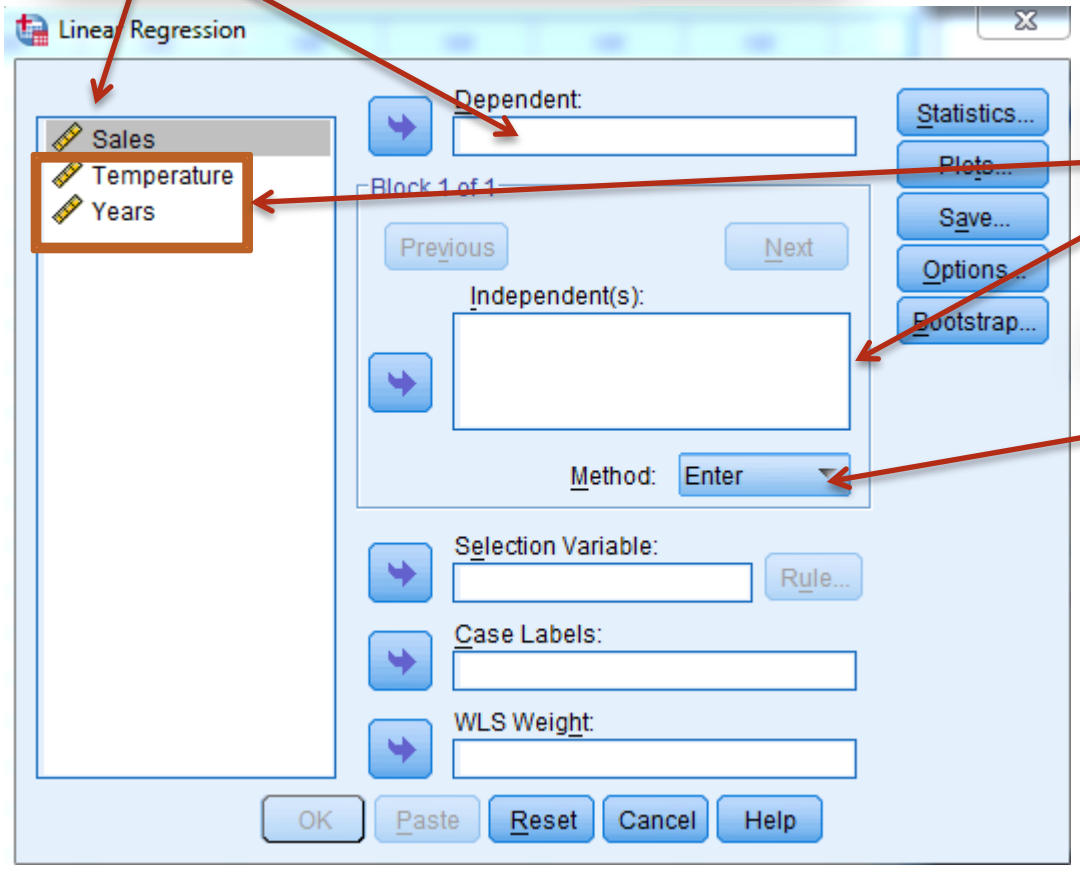
1. الانحدار الخطى المتعدد Multiple Linear Regression
2. الانحدار التدريجى Stepwise Regression
3. الانحدار الهرمى Hierachical Regression
4. الانحدار الغير خطى Curve Linear Regression

الانحدار العياري أو الانحدار الخطي المتعدد
Standard or Simultaneous Regression
Or
Multiple Linear Regression

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



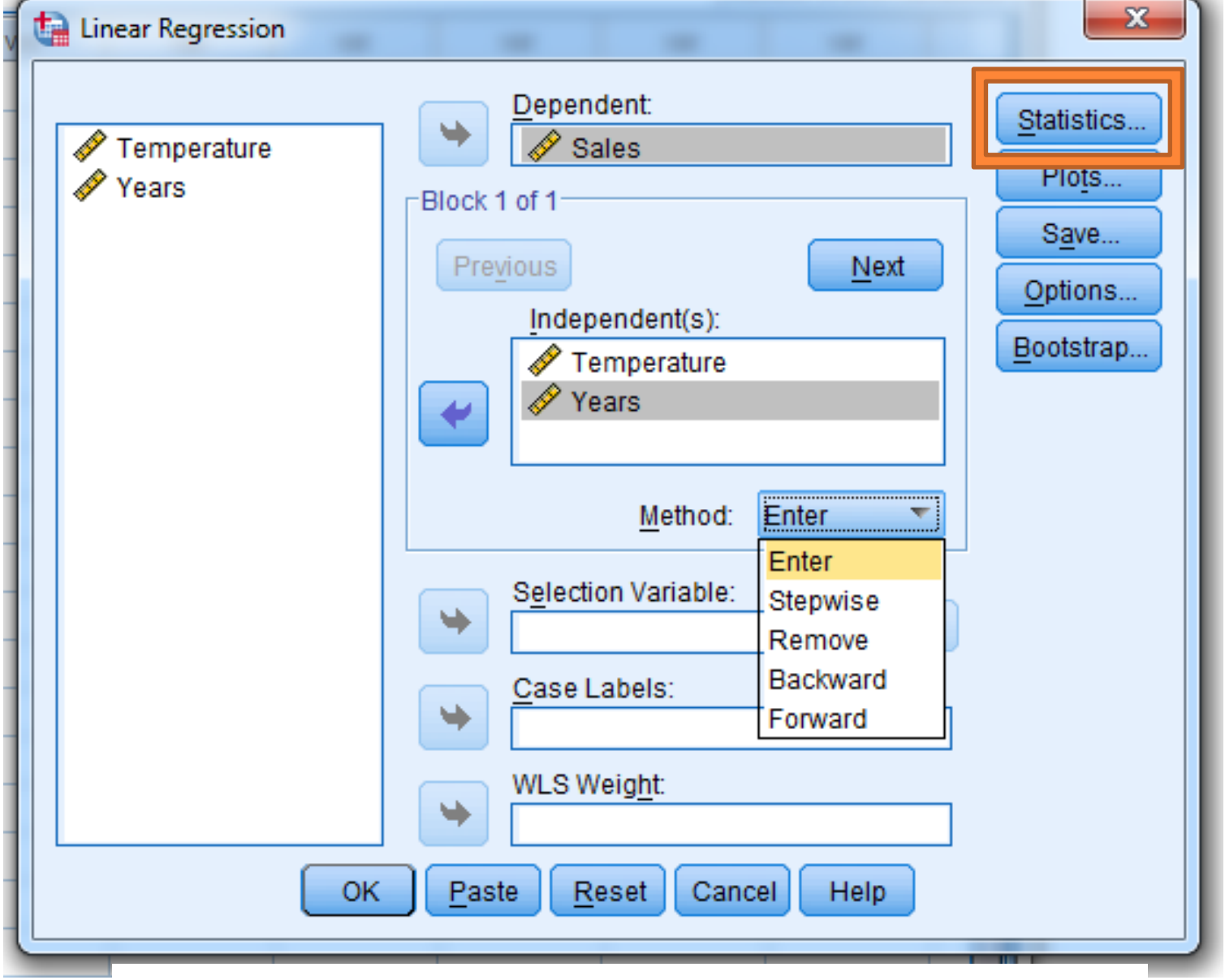
ننقل المتغير التابع Sales لخانة Dependent



ننقل المتغيرات المستقله temp, years
لخانة Independent

نختار نوع الانحدار من خانة Method

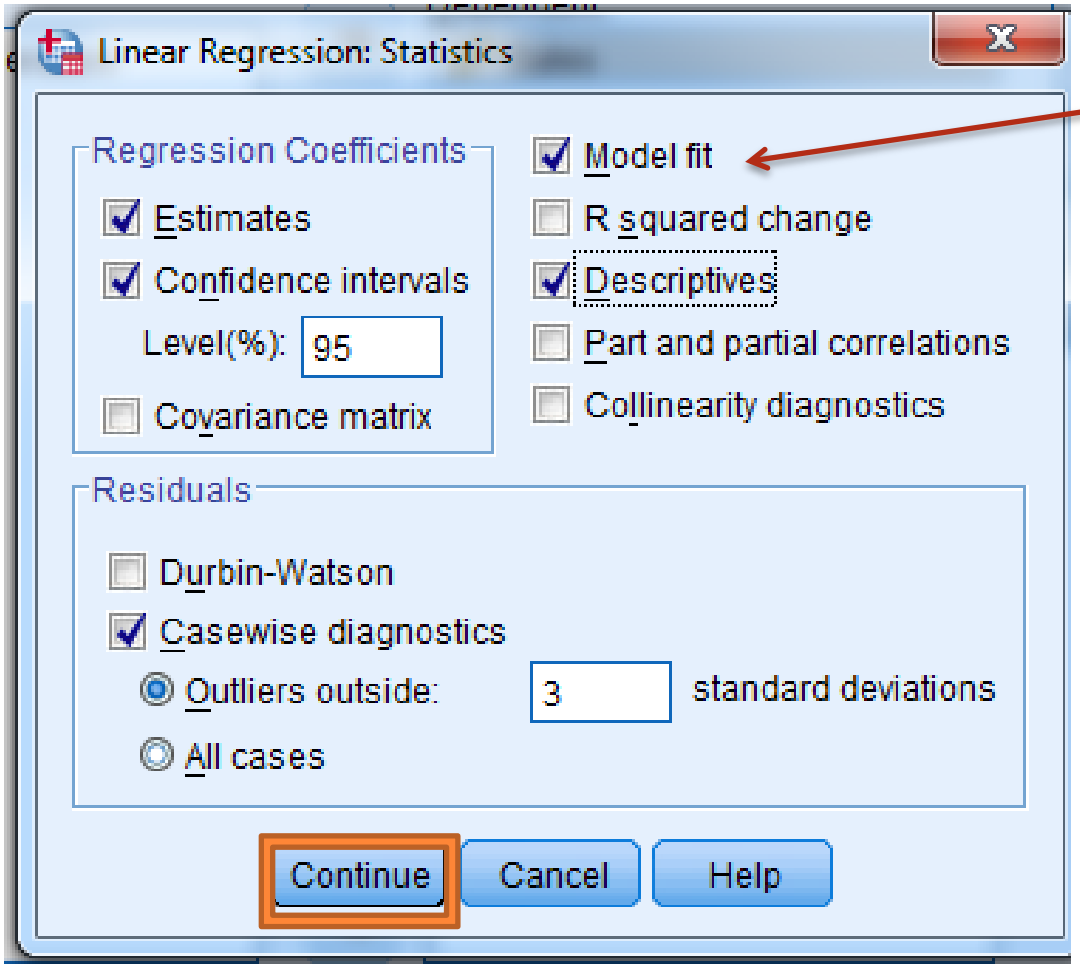
الطريقة العيارية وهى Enter



نضغط على الامر Statistics تظهر شاشه جديده بعنوان Linear Regression: Statistics

نتأكد أن Model fit, Estimates, Casewise diagnostics

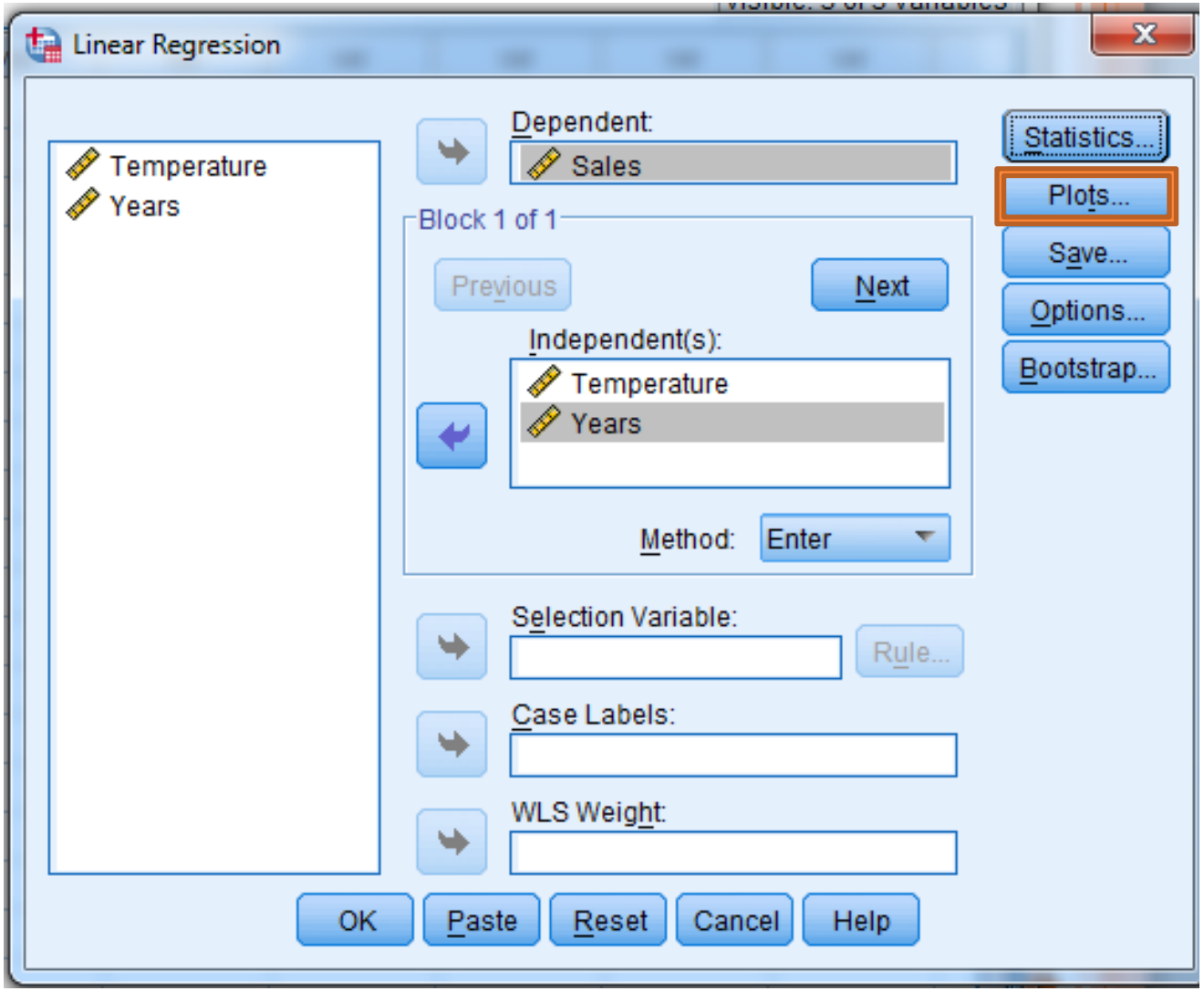
مختاره ويمكن ايضا اضافه اختيارات اخرى



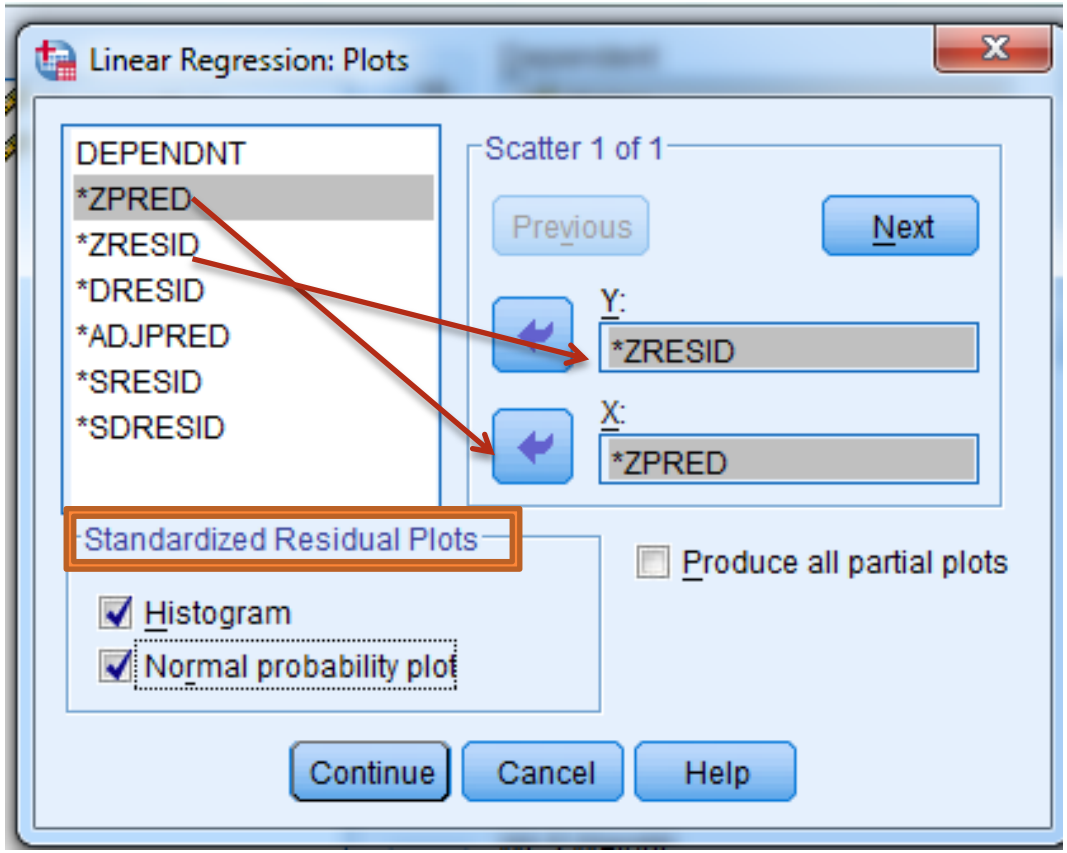
موائمة النموذج (Model fit)

حيث يتم عرض معاملات او مؤشرات حسن المطابقة او الموائمة **Goodness of Fit** للمتغيرات التي يتم ادخالها او اخراجها من النموذج ، وتعرض المؤشرات : معامل الارتباط المتعدد **R** ، ومعامل التحديد R^2 ، وجدول تحليل التباين .

نختار Continue لنعود للشاشة السابقة



نضغط على الامر Plot فتظهر شاشه جديده بعنوان Linear Regression: Plots



الرسوم البيانية تساعد في اختبار صدق الافتراضات الأساسية مثل الإعتدالية والخطية وتجانس التباين ، وكذلك تفيد في الكشف عن الحالات أو الدرجات المتطرفة في التوزيع .

وبالنافذة السابقة يوجد ما يلي :

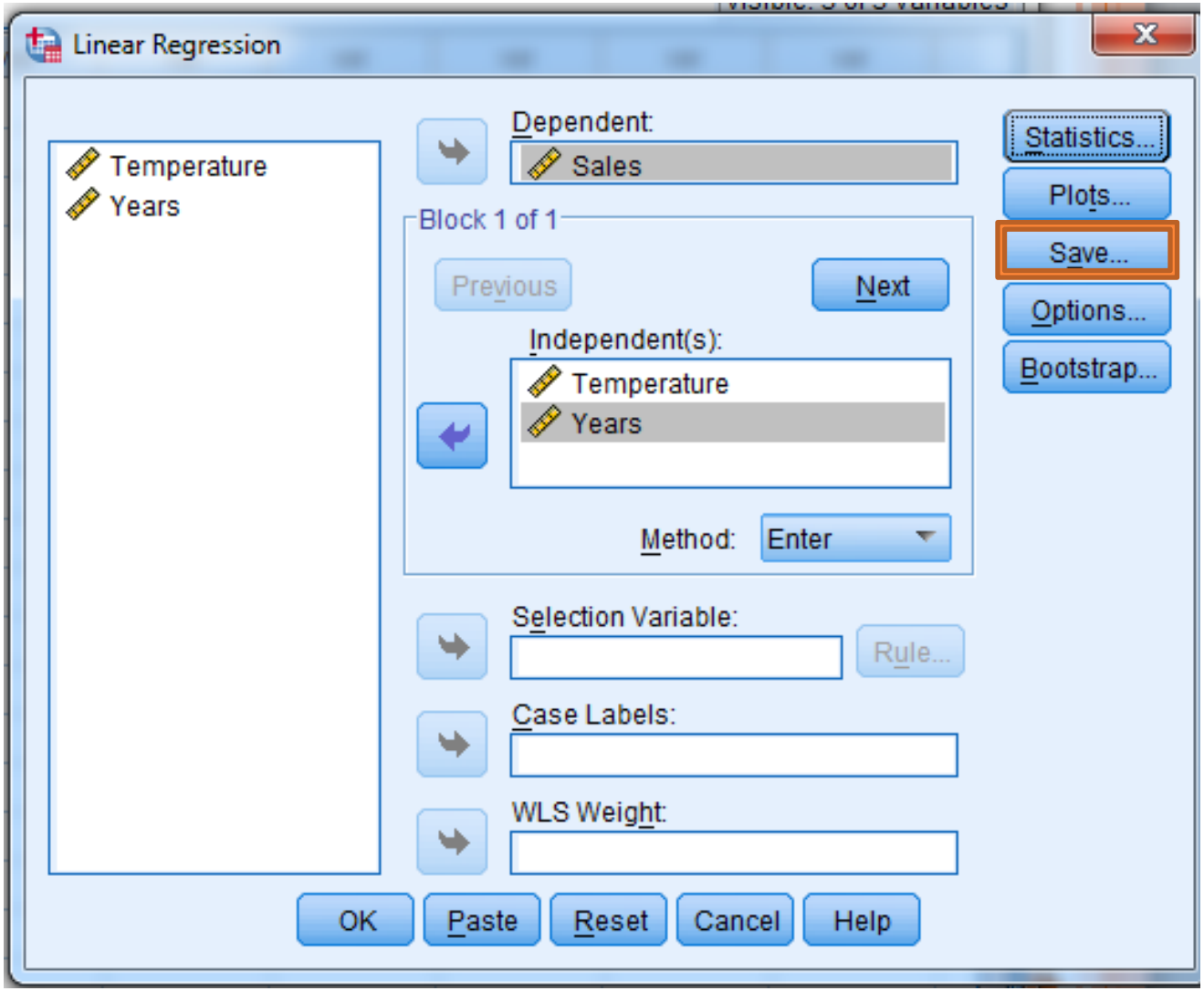
رسم بياني لمخطط شكل الانتشار – **Scatter plots**

حيث يسمح بتحديد أكثر من رسم لمخططات الانتشار في الخانات **X , Y** ويمكن التحول بينهما بالزرين **Previous , next** ويمكننا من رسم شكل الانتشار بين أي اثنين من الآتي : المتغير التابع ، القيم المنبئة **Predicted Values**، البواقي ، البواقي المحذوفة ، القيم المنبئة المعدلة **Adjusted** ، البواقي المعيارية **Standardized** ويمكنك عمل رسم لشكل الانتشار بين البواقي مع القيم المنبئة لاختبار الخطية وتجانس التباين

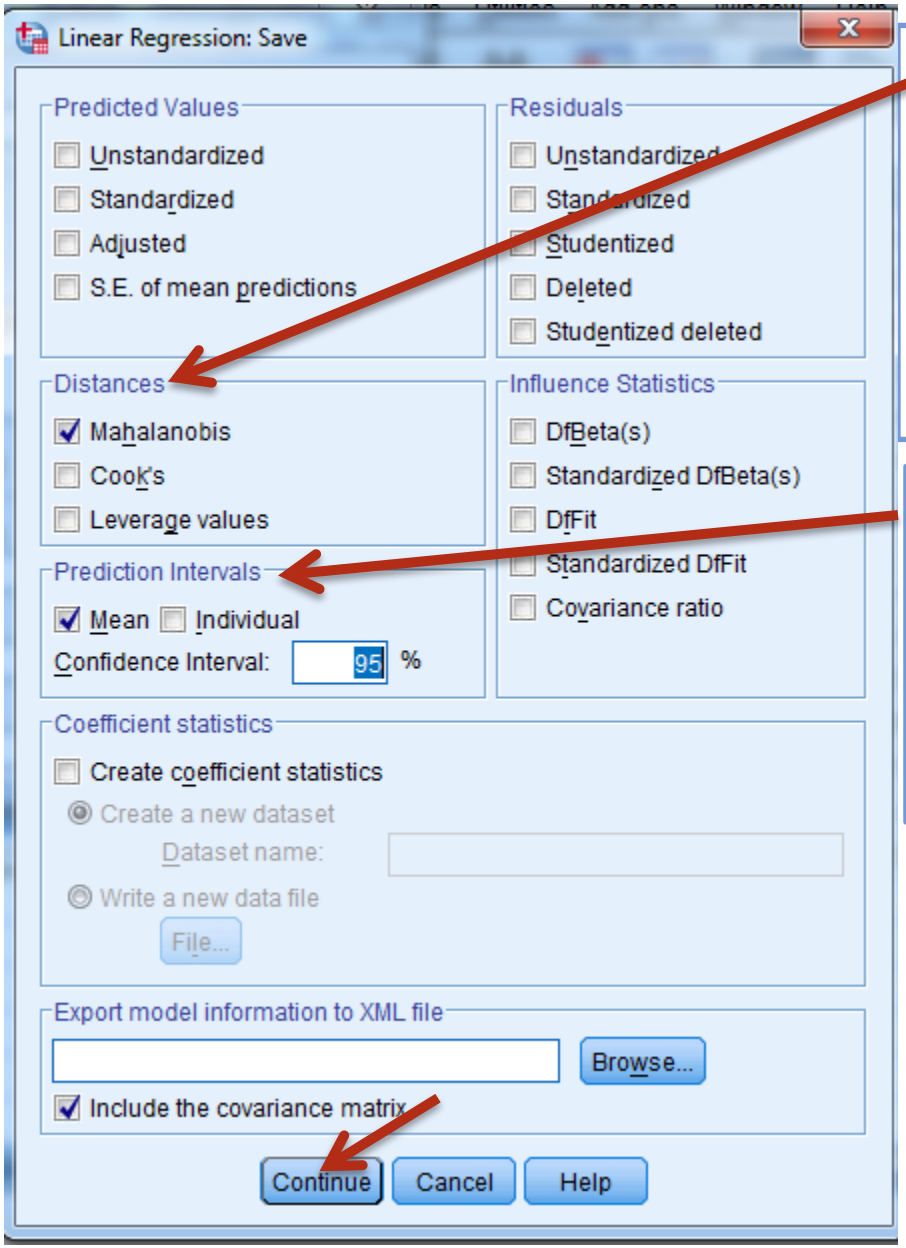
ننقل *ZRESID للمستطيل المقابل لـ Y: وايضا *ZPRED للمستطيل المقابل X:

من قائمة Standardized Residual Plots نختار كلا من Histogram, Normal probability plot

نضغط على Continue فنعود للشاشة السابقة



نضغط على الامر Save تظهر شاشة جديده بعنوان Linear Regression: Save



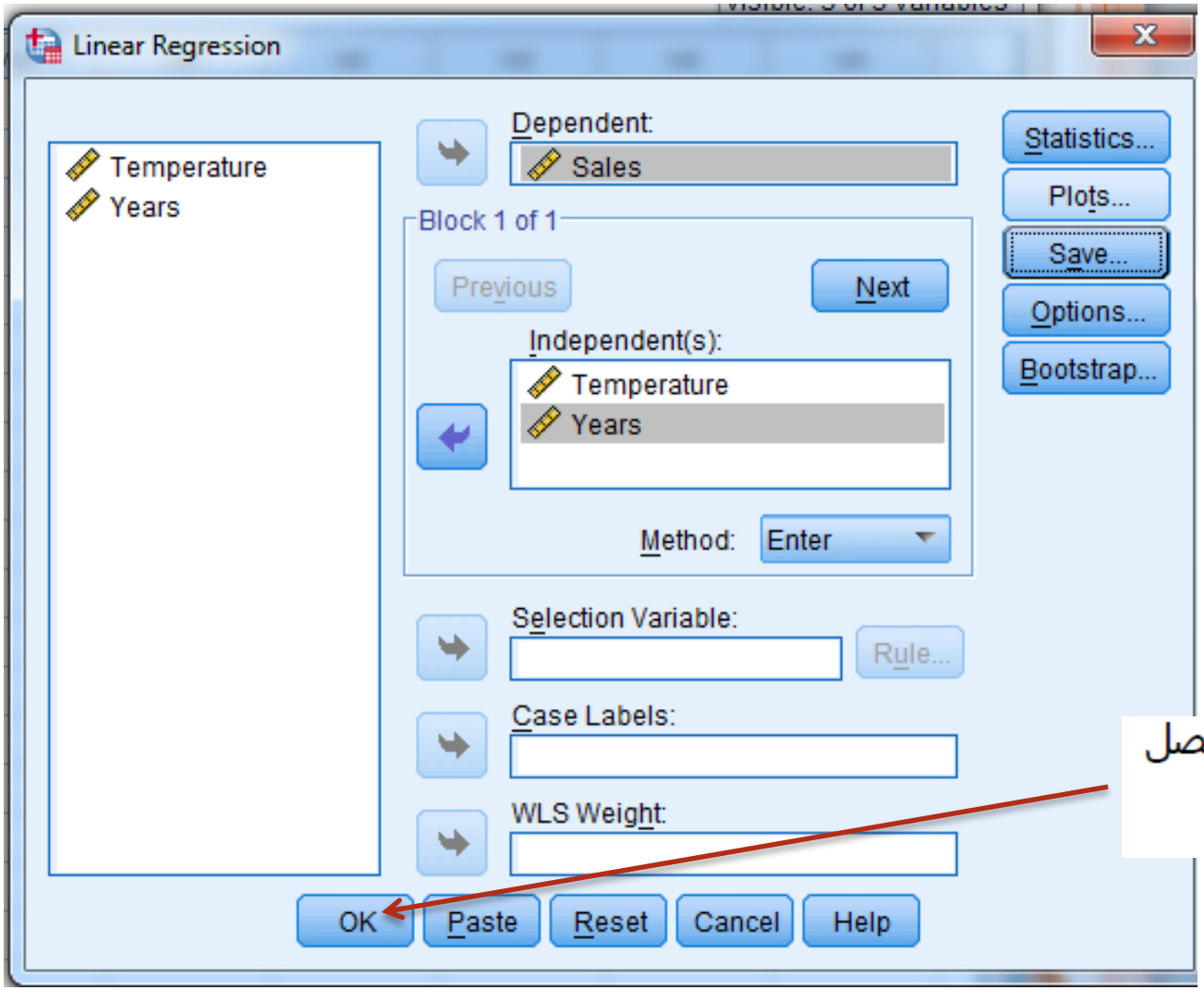
* Distances

وهي مقاييس تهدف إلى تحديد القيم أو الدرجات التي تشمل على تجمعات غير عادية للمتغير المستقل وكذلك الدرجات التي لها تأثير كبير على نموذج الانحدار ، وتقدم الاختبارات لتحديد ذلك وتشمل **Mahalanobis** **Cooks , Leverage** الفئات التنبؤية – **Prediction Intervals**

Prediction intervals

الفئات التنبؤية

وتقدم الحدود العليا والدنيا لكل من فئات المتوسط والدرجات الفردية المنبئة وتمكن المستخدم من تحديد مستوى الدلالة أو الثقة Confidence Interval .



نضغط على Ok فنحصل على النتائج التاليه.

الجدول الأول:

بعنوان Descriptive Statistics ويعطى لنا المتوسط والانحراف المعيارى وعدد الحالات لكل متغير على حده.

Descriptive Statistics

	Mean	Std. Deviation	N
Sales	37.27	14.499	15
Temperature	30.33	8.217	15
Years	2.93	1.580	15

الجدول الثانى:

بعنوان Correlations وهى مصفوفه الارتباط بين جميع المتغيرات وايضا معنويه الارتباط ونلاحظ انه لا يوجد ارتباط تام بين المتغيرات المستقله وبعضها الاخر.

Correlations

		Sales	Temperature	Years
Pearson Correlation	Sales	1.000	.907	.849
	Temperature	.907	1.000	.888
	Years	.849	.888	1.000
Sig. (1-tailed)	Sales	.	.000	.000
	Temperature	.000	.	.000
	Years	.000	.000	.
N	Sales	15	15	15
	Temperature	15	15	15
	Years	15	15	15

Variables Entered/Removed[a]

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

a. Dependent Variable: Sales

b. All requested variables entered.

الجدول الثالث

بعنوان Variables Entered/Removed ويحتوى على أسماء المتغيرات التى دخلت فى معادله الانحدار وهما متغيرين year, temp والمتغيرات التى استبعدت من الدخول فى المعادلة وهنا فى الطريقه العياريه لا تستبعد متغيرات.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.912[a]	.832	.804	6.426

a. Predictors: (Constant), Years, Temperature

b. Dependent Variable: Sales

الجدول الرابع:

بعنوان Model Summary يحتوى على بعض المقاييس التى تم حسابها للنموذج المقدر وهى

1. قيمة معامل الارتباط R=0.912 وهو عالى جدا
2. مربع معامل الارتباط ويستخدم تعين مدى البيانات المستخدمة من المتغيرات المستقله فى تقدير المتغير التابع ونلاحظ أن النموذج المقدر يعبر عن (المتغيرين المستقلين معا) 80% من البيانات وزيادة قيمه هذا المقياس يفسر أن النموذج المقترح ملائم.
3. تعيين مربع معامل الارتباط المعدل Adjusted R Square ويستخدم لنفس الغرض السابق ولكنه أدق.
4. تعيين خطأ التقدير Std. Error of the Estimate وهو هنا 6.42616 كلما قل دل على خطأ أقل للنموذج.

ANOVA[a]

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

a. Dependent Variable: Sales

b. Predictors: (Constant), Years, Temperature

الجدول الخامس:

يحتوى على نتائج تحليل التباين ANOVA لاختبار معنويه الانحدار

الفرض الصفرى: الانحدار غير معنوى

الفرض البديل: الانحدار معنوى

ومن جدول ANOVA نجد ان Sig. = 000 وهى أقل من مستوى المعنويه 0.05 لذا سوف نرفض فرض العدم ونقبل الفرض البديل وهى أن الانحدار معنوى وبالتالى توجد علاقة ما بين المتغيرات المستقله والمتغير التابع.

النموذج ككل معنوي بدون تحديد أي من المعاملات هو السبب في المعنوية

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-6.993-	8.428		-.830-	.423	-25.356-	11.371
	Temperature	1.275	.454	.722	2.807	.016	.285	2.264
	Years	1.907	2.362	.208	.807	.435	-3.240-	7.054

a. Dependent Variable: Sales

الجدول السادس:

بعنوان Coefficients ويساعد هذا الجدول فى الحصول على كلا من

1. معادلة خط الانحدار المقدرة والخطأ فى التقدير لكل معامل وذلك من العمود Unstandarized Coefficients حيث

$$Sales = -6.993 + 1.275\,temp + 1.907\,years$$

قيمة sig اكبر من 0.05 لذا فان الثابت ومعامل السنوات غير معنوى لكن فى حاله درجات الحرارة sig. =0.016 أقل من 0.05 لذا فان معامل درجات الحرارة معنوى وهو سبب معنوية تحليل التباين للانحدار.

العمود الاخير يقدم التقدير بفترة لمعاملات خط الانحدار والثابت

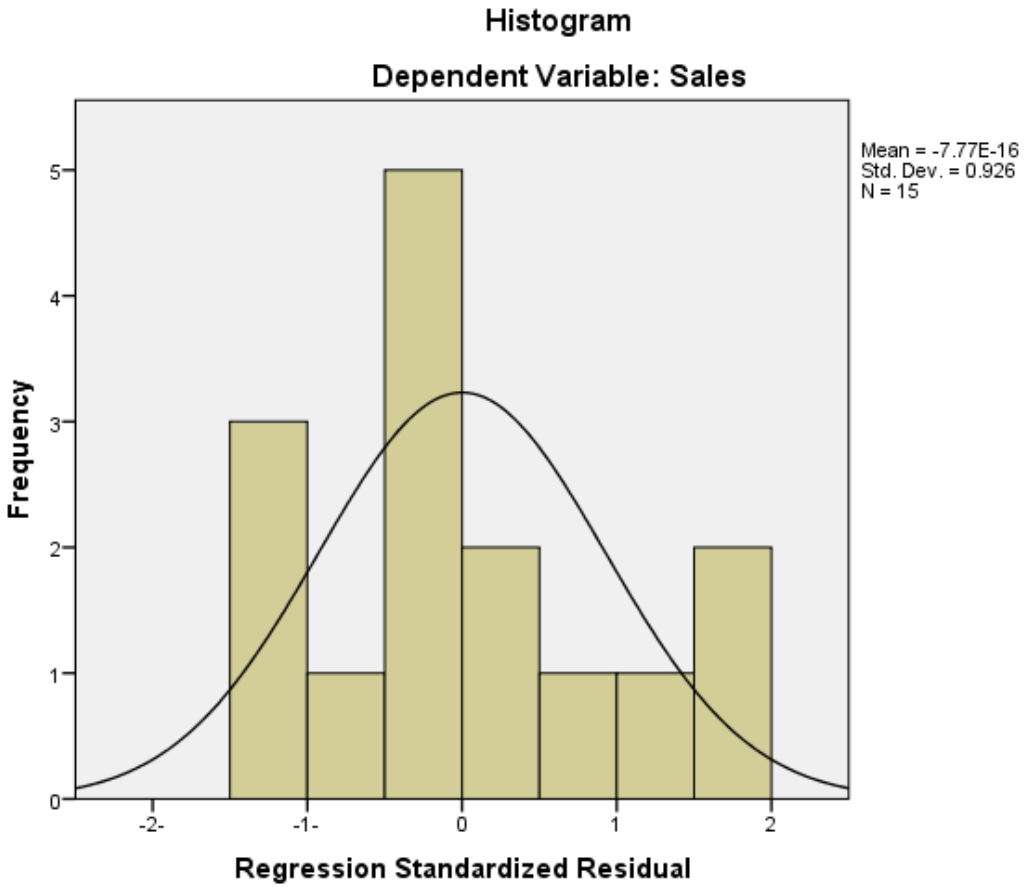
Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	17.86	55.44	37.27	13.222	15
Std. Predicted Value	-1.468-	1.374	.000	1.000	15
Standard Error of Predicted Value	1.945	4.247	2.801	.667	15
Adjusted Predicted Value	18.72	59.65	37.52	13.420	15
Residual	-7.255-	10.294	.000	5.949	15
Std. Residual	-1.129-	1.602	.000	.926	15
Stud. Residual	-1.257-	1.826	-.017-	1.034	15
Deleted Residual	-9.654-	13.374	-.254-	7.483	15
Stud. Deleted Residual	-1.291-	2.057	.008	1.090	15
Mahal. Distance	.349	5.182	1.867	1.347	15
Cook's Distance	.000	.333	.090	.113	15
Centered Leverage Value	.025	.370	.133	.096	15

a. Dependent Variable: Sales

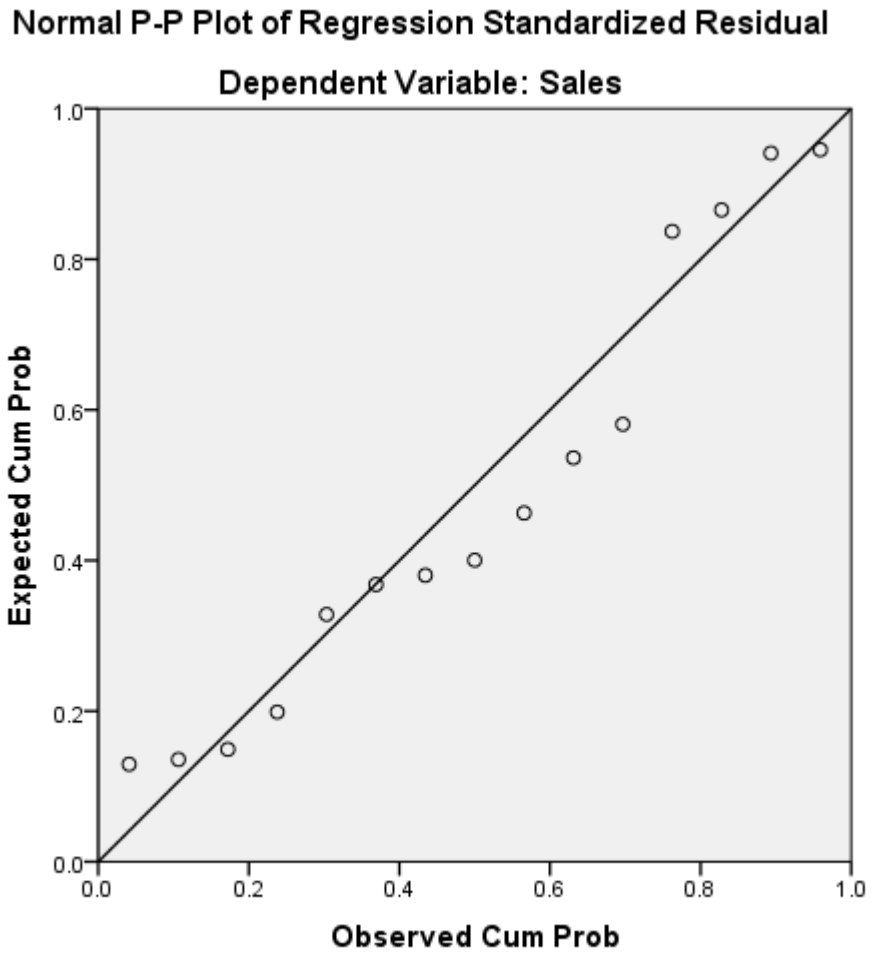
الجدول الثامن:

بعنوان Residuals Statistics يستخدم لمعرفه بعض المقاييس الخاصه بالبواقى



الشكل البيانى:

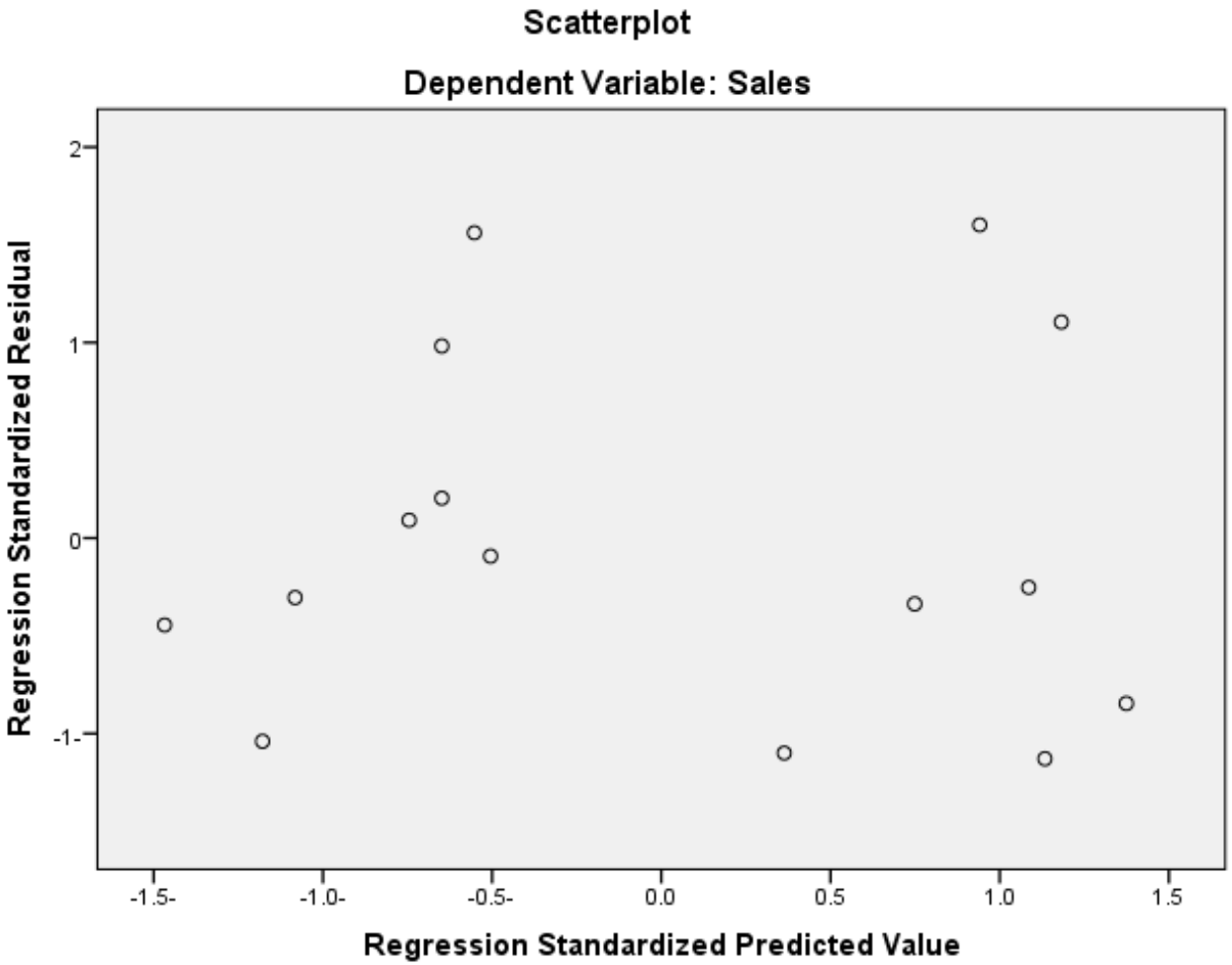
هو المدرج التكرارى ويستخدم للتعرف هل البيانات تتوزع حسب التوزيع الطبيعى أم لا؟



الشكل البيانى:

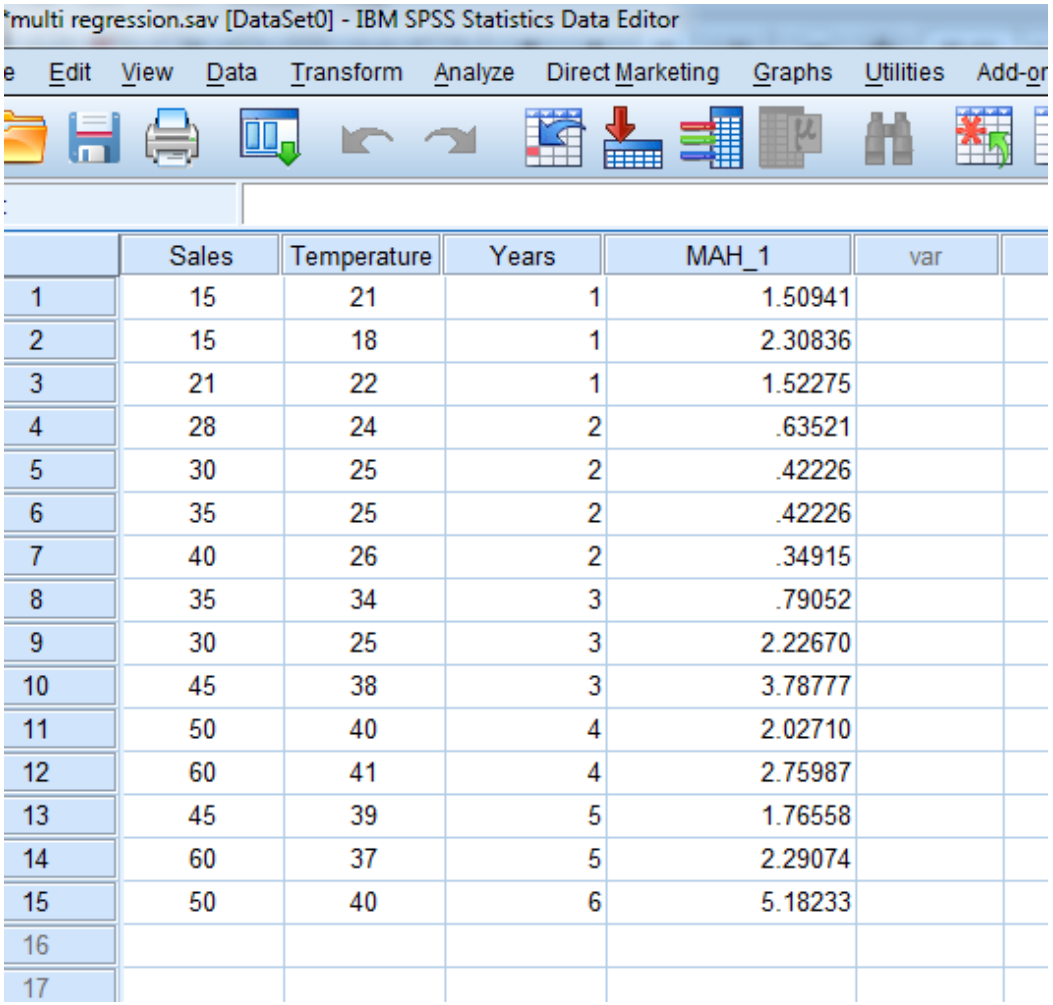
يختبر هل البواقى تتبع التوزيع الطبيعى ام لا؟

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



الشكل البياني:

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



	Sales	Temperature	Years	MAH_1	var
1	15	21	1	1.50941	
2	15	18	1	2.30836	
3	21	22	1	1.52275	
4	28	24	2	.63521	
5	30	25	2	.42226	
6	35	25	2	.42226	
7	40	26	2	.34915	
8	35	34	3	.79052	
9	30	25	3	2.22670	
10	45	38	3	3.78777	
11	50	40	4	2.02710	
12	60	41	4	2.75987	
13	45	39	5	1.76558	
14	60	37	5	2.29074	
15	50	40	6	5.18233	
16					
17					

بالعودة لملف البيانات نجد انه قد أضيف متغير جديد mah_1 وذلك لأننا طلبنا اختبار Mahalanobis فنقوم بمقارنة قيم هذا المتغير بقمية Chi-Square عند درجة حريه n-1 = 2 ومستوى معنويه مثلا 0.001 فنجد أن chi-Square = 13.8 وجميع قيم المتغير أقل من هذه القيمه لذا فانه لا يوجد قيم متطرفه متعدده.

١- إن ادخال عدد كبير من المتغيرات في النموذج يجعل من الصعب تحقيق شروط تطبيق الانحدار (الارتباط الذاتي والخطية والتجانس،.....

٢- هناك شرط لابد من تحققه في عدد القراءات وعدد المتغيرات (وهو أن عدد القراءات عشرون ضعف لعدد المتغيرات) فإذا كان عدد المتغيرات لم يتحقق ؟

٢- كيف نعالج عدم معنوية بعض المتغيرات.

سنستخدم الانحدار الهرمي (خطوة خطوة) والذي يهدف أساسا إلى إيجاد علاقة بين المتغير التابع والمتغيرات المستقلة الأكثر ارتباطا به ويتم ذلك بشكل تدريجي