الارتباط والانحدار

كيفية استخدم إكسل لحساب معامل الارتباط:

هناك طريقتان يمكننا استخدامهما.

افترض أن لدينا البيانات التالية:

أ	ب	ت	ث
1	25	3	22
2	39	37	40
3	35	3	30
4	30	15	43
5	39	10	30
6	27	13	44
7	29	22	36
8	33	16	50
9	32	28	38
10	31	15	30
11	32	34	22
12	34	22	30
13	28	46	35
14	27	43	50
15	15	47	40

ونريد حساب معامل الارتباط بين المتغير أ و ب. الطريقة الأولى هي أن نستخدم الدالة المتاحة في إكسل لحساب معامل الارتباط فنكتب ما يلي في أي خلية:

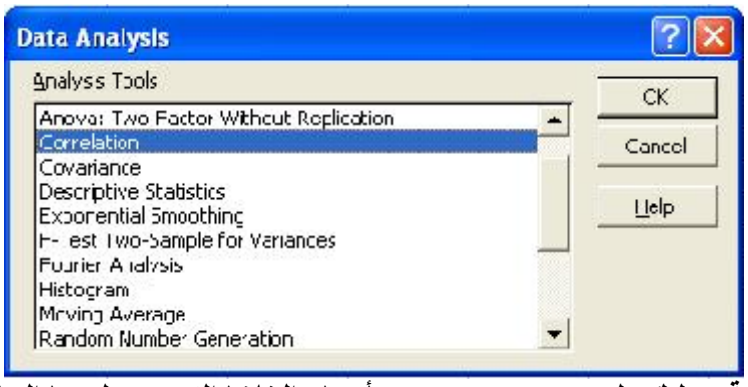
=CORREL(D2:D16,C2:C16)

وبالتالي نحصل على معامل الارتباط بين أ و ب وهو - ٠.٤٦ . هذه القيمة تعني وجود علاقة عكسية ضعيفة لأن القيمة لا تقترب من- ١ بل هي أقرب قليلا إلى الصفر.

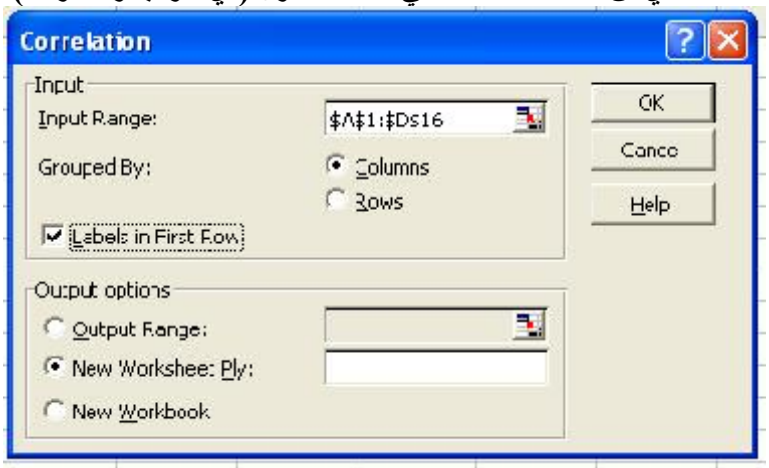
الطريقة الثانية تساعدنا في الحصول على معامل الارتباط بين متغيرين أو عدة متغيرات مرة واحدة. هذه الطريقة تتم كالتالي:

اضغط على Tools ثم Data Analysis)

ثم اختر Correlation



تظهر لك النافذة التالية وعليك ملء Input Range بأسماء الخلايا التي مسجل بها البيانات. وقد علمت على Labels in First Row أي أن أسماء الأعمدة في الصف الأول (أي أ و ب و ت و ث)



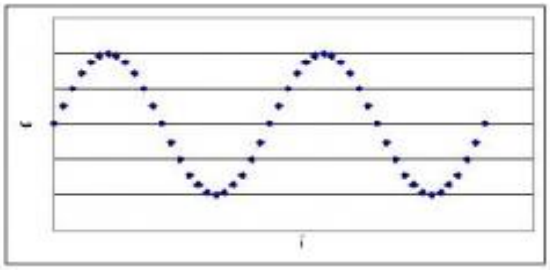
نضغط OK فنحصل على النتيجة كالتالي:

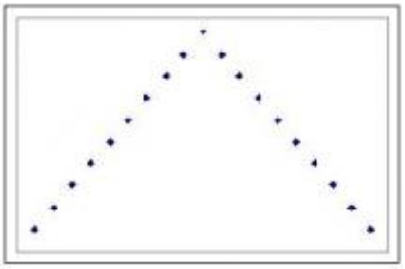
	ث	ت	ب	أ
ث	1			
ت	0.21	1		
ب	0.12	-0.35	1	
أ	0.25	0.72	-0.46	1

هذا الجدول (بالأعلى) يبين أن معامل الارتباط بين ث و ت مثلا هي ٠.٢١ ومعامل الارتباط بين ب و ت هي ٠.٣٥- وهكذا. بالطبع فإن العلاقة بين المتغير ونفسه هي ١ فترى في الجدول معامل ارتباط ث بـ ث هو ١ وهي قيمة لا تعنينا في شيء. هذه الطريقة سريعة جدا عندما يكون لدينا أكثر من متغيرين. من هذه النتيجة نرى أن العلاقة الخطية بين أ و ت هي الوحيدة التي يمكن أخذها في الاعتبار لأنها تساوي ٠.٧٢ أما باقي القيم فهي صغيرة جدا.

هل لا توجد علاقة؟

ليس معنى أن يكون معامل الارتباط صفرا أو قريبا من الصفر أنه لا توجد أي علاقة بين المتغيرين. فمعامل الارتباط يبين قوة العلاقة الخطية. والعلاقة الخطية هي علاقة في شكل خط مستقيم فهي علاقة ليس بها منحنيات أو طلوع ونزول. فالعلاقة الخطية تكون طردية أو عكسية فقط. وبالتالي فقد يكون معامل الارتباط يساوي صفرا ولكن توجد علاقة قوية بين المتغيرين ولكنها غير خطية أي أنها ليست على شكل خط مستقيم كما في الامثلة التالية:





ففي هذين الشكلين نرى علاقة واضحة بين المتغيرين ولكنها ليست مجرد علاقة طردية أو عكسية ولا يمكن تمثيلها بخط مستقيم. ففي الحالة الأولى نلاحظ تغير المتغير الثاني بشكل دوري مع المتغير الأول. وفي الحالة الثانية نجد علاقة طردية حتى نقطة ما ثم تتحول العلاقة إلى علاقة عكسية. هذه العلاقات هي علاقات غير خطية ولا يمكن التنبؤ بها بمعامل الارتباط.

بهذا نكون قد استطعنا دراسة شكل العلاقة عن طريق منحنى الانتشار (المنحنى التنقيطي) ومعرفة قوة العلاقة الخطية عن طريق معامل الارتباط.

مثال١: أوجد معامل الارتباط بين دخل تسع أسر (X) والإنفاق (Y) الشهري بالالف ريالوالمبينة في الجدول الآتي:

X	6	8	7	14	11	12	8	9	10
Y	4	8	6	10	9	11	8	7	8

نكون جدول البيانات حسب احد القوانين المذكورة أعلاه وليكن الثاني مع العلم بأن الصف الخير يعبر عن المجموع للبيانات في العمود والتي تقع أعلاه.

X	Y	XY	X^2	Y^2
6	4	24	36	16
8	8	64	64	64
7	6	42	49	36
14	10	140	196	100
11	9	99	121	81
12	11	132	144	121
8	8	64	64	64
9	7	63	81	49
10	8	80	100	64
85	71	708	855	595

$$r = \frac{n(\sum X_i Y_i) - (\sum X_i)(\sum Y_i)}{\sqrt{n(\sum X_i^2) - (\sum X_i)^2}\sqrt{n(\sum Y_i^2) - (\sum Y_i)^2}}$$

$$r = \frac{9 \times 708 - 85 \times 71}{\sqrt{9 \times 855 - (85)^2}\sqrt{9 \times 595 - (71)^2}}$$

$$r = \frac{6372 - 6035}{\sqrt{7695 - 7225}\sqrt{5355 - 5041}}$$

$$r = \frac{337}{\sqrt{470}\sqrt{314}}$$

$$r = 0.877$$

ملاحظة:
قيمة معامل الارتباط ٠.٨٧٧ يعني قوة الارتباط قوية (من ٠.٧ إلى ٠.٩٩) حيث أن قيمة r تحدد قوة الارتباط وهذا بيان بقوة الارتباط من صفر إلى ١:
الاختبار هنا: H_o : $\rho = 0$ أي لا توجد علاقة .
H_1 : $\rho \neq 0$ أي توجد علاقة.

$$T = r\sqrt{\frac{n-2}{1-r^2}}$$
$$T = 0.877\sqrt{\frac{9-2}{1-0.769}}$$
$$T = 0.877\sqrt{30.30}$$
$$T = 0.877 \times 5.51$$
$$T = 4.832$$

إحصاء الاختبار المتبع هنا هو:
نرفض H_o بمستوى معنوية ٠.٠٥ إذا كان $T > t_{0.05,7}$ أي إذا كان $T > ١.٨٣٣$ وحيث أن $T = 4.832$ وهي أكبر من ١.٨٣٣ فنرفض H_o ونقبل H_1 بوجود علاقة أي نرفض الفرضية الصفرية ونقبل بالفرضية البديلة أي توجد علاقة.

معامل ارتباط فاي (Coefficient Phi Correlation)

إن الحالة التي تكون فيها البيانات للمتغيرين x , y غير قابله للترتيب التصاعدي أو التنازلي كمتغير الجنس (ذكر ، أنثى) أو متغير التدخين (مدخن ، غير مدخن) أو ... وعليه نكون جدول نواتجه في ٢ × ٢ خانات للمتغيرين والصفتين يكتب ببساطة على الصورة:

	1	2
a	n_{a1}	n_{a2}
b	n_{b1}	n_{b2}

ويمكن صياغة الجدول بالصورة التالية لتسهيل الحسابات:

	x_i	x_i	Total
y_i	a	b	a + b
y_i	c	d	c d
Total	a + c	b + d	a − b − c + d

ويحسب معامل الاقتران من الصيغة الرياضية:

$$r_\phi = \frac{bc - ad}{\sqrt{(a+c)(b+d)(c+d)(a+b)}}$$

مثال:

في دراسة على ٢٥ شخص لمعرفة العلاقة بين الجنس والتدخين فجمعت البيانات في الجدول التالي والمطلوب حساب معامل ارتباط فاي.

	ذكر	أنثى
يدخن	10	2
لا يدخن	8	5

الحل:

نكمل الجدول بعملية الجمع بالصورة الآتية:

	أنثى	ذكر	المجموع
يدخن	2	10	12
لا يدخن	5	8	13
المجموع	7	18	25

وبتطبيق القانون نجد أن:

$$r_\phi = \frac{bc - ad}{\sqrt{(a+c)(b+d)(c+d)(a+b)}}$$

$$r_\phi = \frac{10 \times 5 - 2 \times 8}{\sqrt{7 \times 18 \times 13 \times 12}}$$

$$r_\phi = \frac{34}{\sqrt{19656}}$$

$$r_\phi = 0.243$$

القيمة الموجبة هنا لمعامل فاي تبين بأن الذكور أكثر ميلاً للتدخين من الإناث كما أن قيمة المعامل هنا ٠.٢٤٣ تدل على ضعف العلاقة.

أمثلة إضافية غير محلولة

(١)اذاكان لديك الاستبانة التالية وفيها :

١ دخل الأسرة .....

٢ قيمة الأستهلاك الشهري.....

اجابات الدخل	٤٠٠٠	٣٥٠٠	٢٢٠٠	٣٢٠٠	٢٥٠٠	٢٧٠٠	٢٨٠٠	٣٣٠٠	٢٦٠٠	٢٠٠٠
اجابات الاستهلاك	٣٢٠٠	٢٧٠٠	١٧٠٠	٢٠٠٠	١٨٠٠	١٩٠٠	٢٦٠٠	٣٠٠٠	٢٢٠٠	٢٠٠٠

حدد قيمة معامل الارتباط مع تحديد نوعه وقوته

اختبر الفرضية القائلة بأنه لايوجد ارتباط ذو دلالة معنوية بين كلا من الدخل والاستهلاك مستخدما دلالة ٠.٠٥

(٢)اذاكان لديك الاستبانة التالية وفيها :

١ دخل الأسرة .....

٢ قيمة الأستهلاك الشهري.....

٣ عدد افراد الأسرة .....

اجابات الدخل	٤٠٠٠	٣٥٠٠	٢٢٠٠	٣٢٠٠	٢٥٠٠	٢٧٠٠	٢٨٠٠	٣٣٠٠	٢٦٠٠	٢٠٠٠
اجابات الاستهلاك	٣٢٠٠	٢٧٠٠	١٧٠٠	٢٠٠٠	١٨٠٠	١٩٠٠	٢٦٠٠	٣٠٠٠	٢٢٠٠	٢٠٠٠
اجابات عدد الاسرة	٧	٨	٥	٣	٤	٣	٥	٣	٢	٢

حدد قيم معاملات الارتباط مع تحديد النوع والقوة

هل تدل البيانات على عدم وجود ارتباط ذو دلالة معنوية بين المتغيرات مستخدما دلالة ٠.٠٥

(٣)اذا كانت لديك الاستبانة التالية وفيها :

المؤهل العلمي : ١ دبلوم ٢ بكالوريوس ٣ ماجستير ٤ دكتوارة

الحالة الاجتماعية : ١ أعزب ٢ متزوج ٣ أرمل ٤ مطلق

المؤهل العلمي	٤	٣	١	٢	٤	٢	٤	١	١	٢
الحالة الاجتماعية	٢	١	٤	٢	٢	٣	٢	١	٢	٣

حدد قيم معاملات الارتباط مع تحديد النوع والقوة

هل تدل البيانات على عدم وجود ارتباط ذو دلالة معنوية بين المتغيرات مستخدما دلالة ٠.٠٥

ثانياً : تحليل الانحدار

تحليل الانحدار Analysis Of Regression

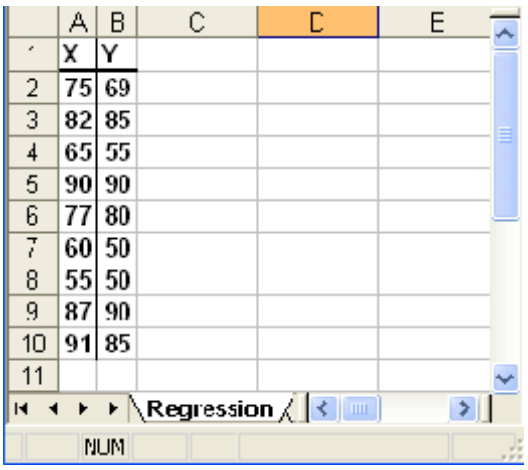
- **الانحدار الخطي البسيط Simple Linear Regression :**

الشكل العام لمعادلة الانحدار الخطي البسيط : $Y = \beta_0 + \beta_1 X + \varepsilon$

نرمز للمعادلة المقدرة : $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X + e$ حيث e يسمى الراسب (أو الباقي أو الخطأ) : $e = Y - \hat{Y}$

❖ لإجراء تحليل الانحدار الخطي البسيط في برنامج إكسل نقوم بالتالي:

١. لتكن لدينا البيانات الموضحة في الشكل المجاور:

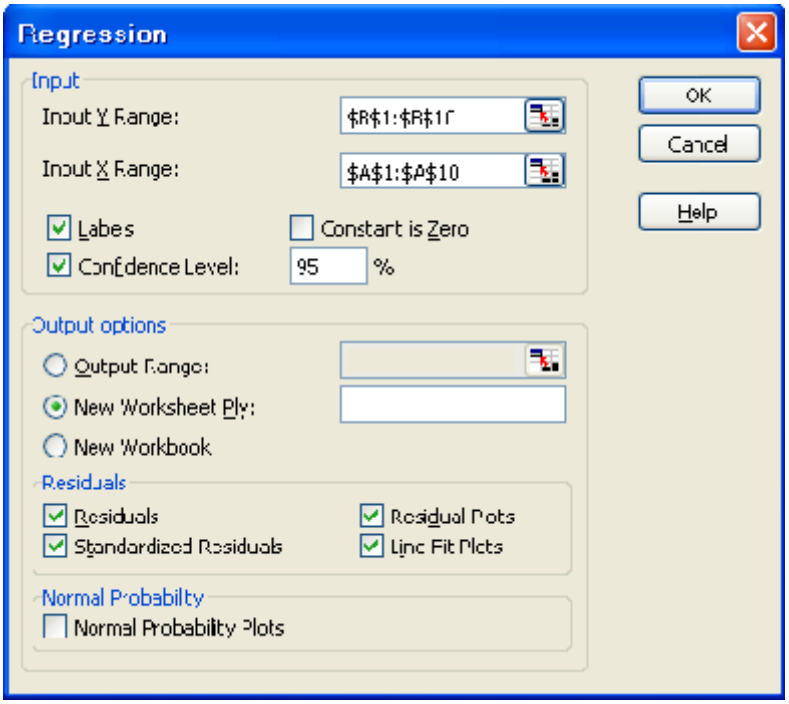


	A	B	C	D	E
1	X	Y			
2	75	69			
3	82	85			
4	65	55			
5	90	90			
6	77	80			
7	60	50			
8	55	50			
9	87	90			
10	91	85			
11					

Regression

NUM

من شريط القوائم : أدوات (Tools) > Data Analysis > Regression : فتظهر لدينا النافذة التالية :



في حقل Input Y Range نكتب نطاق خلايا قيم المتغير التابع Y (Dependent Variable)
في حقل Input Y Range نكتب نطاق خلايا قيم المتغير المستقل (Independent Variable) X
كما ترى في الشكل فإن تم تفعيل الخيار Labels ، ذلك لأننا كتبنا المتغيرات(x و y) في الخلايا الأولى من الأعمدة وعليه يجب تفعيل هذا الخيار لإعلام البرنامج بذلك
٢. إذا تم تفعيل الخيار Confidence Level فسيتم إظهار فترات الثقة للمعاملات المقدرة
بالنسبة للبواقي :
بتفعيل الخيار Residuals سيتم إظهار جدول البواقي، وبتفعيل الخيار Standardized Residuals سيتم إظهار جدول للبواقي المعايرة (سيتم توضيحهم لاحقا)
بتفعيل الخيار Residual Plots سيتم تمثيل البواقي بيانيا ،
بتفعيل الخيار Line Fit Plots سيتم رسم لوحة الانتشار للمتغيرين X,Y ، إضافة إلى النقاط المقدرة (خط الانحدار).
بعد أن تتأكد من المدخلات قم بالنقر على OK ،لتظهر لديك أربعة جداول

الجدول الأول: *Regression Statistics*- إحصاءات الانحدار.

SUMMARY
OUTPUT

Regression Statistics	
Multiple R	0.965950001
R Square	0.933059405
Adjusted R Square	0.923496463
Standard Error	4.702076381
Observations	9

القيمة المطلقة لمعامل الارتباط .....
معامل التحديد (وهو مربع معامل الارتباط) ...
يفيد في حالة الارتباط المتعدد .....
الخطأ المعياري للتقدير ،أي هو $\sqrt{MRS}$ ...
عدد المشاهدات n.....

الجدول الثاني: ANOVA – تحليل التباين-

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	2157.233344	2157.233344	97.57032808	2.31997E-05
Residual	7	154.7666561	22.1095223		
Total	8	2312			

نلاحظ أن $P-value(Sign.F)=2.31997E-05=2.31997\times10^{-5}=0.0000231997\prec\alpha=0.05$
وذلك يعني أننا نرفض الفرضية العدم القائلة أنه لا توجد علاقة خطية بين المتحولين X,Y ، ونقبل بالفرضية البديلة أي يوجد علاقة خطية بين المتحولين عند مستوى أهمية alpha = 0.05

الجدول الثالث : جدول المعاملات.

		Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lowe 95%*	*Upper95%*
$\hat{\beta}_1$	Intercept	-21.48020353	9.659200021	-2.22380771	0.061530537	-44.3205	1.360175
$\hat{\beta}_0$	X	1.242407378	0.125778132	9.877769388	2.31997E-05	0.944989	1.539825
		قيم المعاملات المقدرة وفقا لطريقة المربعات الصغرى	الخطأ المعياري للعامل المقدُّر	قيمة مؤشر الاختبار المتعلق بالفرضية	قيمة الاحتمال المتعلقة بالفرضية	حدود مجال الثقة لقيمة كل معامل	

الفرضيات المقصود بها في هذا الجدول هي :

$H_0:\beta_0=0$ مقابل الفرضية البديلة : $H_1:\beta_0\neq0$ وذلك بالنسبة للمعامل $\hat{\beta}_0$ الذي في السطر الثاني .

$H_0:\beta_1=0$ مقابل الفرضية البديلة : $H_1:\beta_1\neq0$ وذلك بالنسبة للمعامل $\hat{\beta}_1$ الذي في السطر الأول.

الجدول الرابع : جدول الراسب و التقديرات:

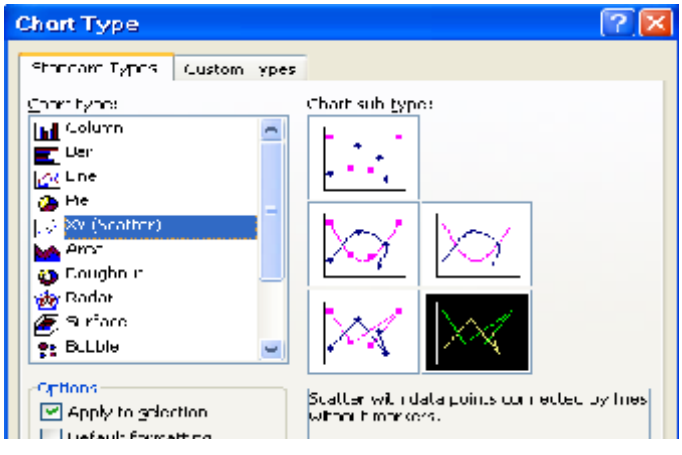
Observation	*Predicted Y*	*Residuals*	*Standard Residuals*
المشاهدة رقم	القيمة المقدرة وفقا لمعادلة الانحدار المقدرة $\hat{Y}$ أي	قيمة الراسب ، $e=Y-\hat{Y}$ أي	الراسب المعاير و تتم معايرته بتقسيمه على الخطأ المعياري للتقدير

أخيرا يتبقى لدينا الرسوم البيانية :

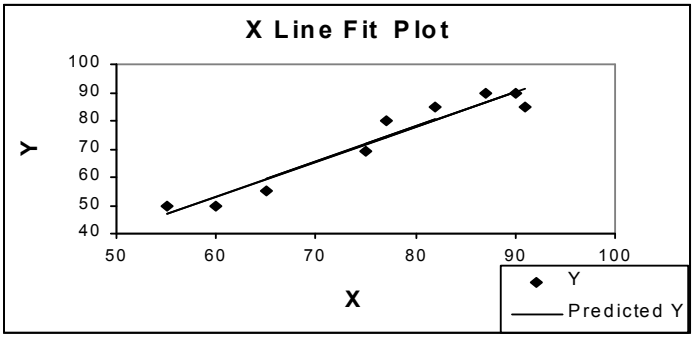
المخطط البياني الأول هو : X Line Fit Plot ، النقط التي باللون النفسجي تمثل النقط: (X_i,Y_i) ، أما النقط باللون الوردي (Predicted Y) فهي تمثل النقط المقدرة وفقا لمعادلة الانحدار ، لاحظ أنها تشكل مستقيما .حتى يكون الشكل مأولفا أكثر أنصحك بالتالي : انقر بالزر الأيمن للفأر فوق إحدى النقاط المقدرة (اللون الوردي) عندها تظهر لديك قائمة مختصرة ، اختر منها " نوع التخطيط " "Chart Type" ستظهر لديك النافذة :

اختر نوع التخطيط المشار إليه (س و ص مبعثر XY Scatter) ، انقر موافق ، ثم إذا أردت أجر بعض التعديلات في الأبعاد و اللون ، وخواص الترقيم (يمكنك أن تبدأ الترقيم من القيمة 50 إذا أردت) يمكنك أن تجري التعديل الذي تريد على أي جزء من التخطيط فقط بالنقر مرتين على ذلك الجزء .

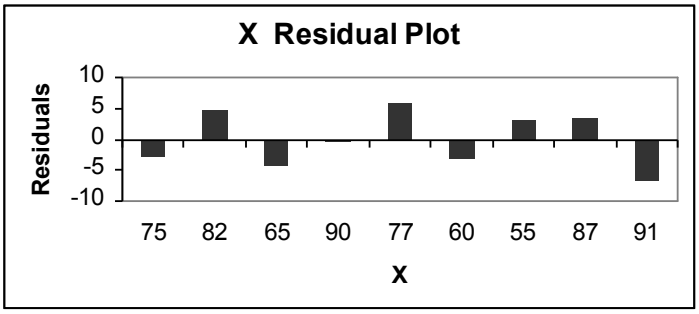
جرب ذلك بنفسك !



سيظهر لديك شكل مشابه للشكل :



الشكل البياني الثاني الذي سيظهر
أيضا هو X Residual Plot ، وهو التمثيل
البياني لقيم الرواسب ، التمثيل يظهر
بشكل نقطي ، يمكنك أيضا التعديل على
نوع التخطيط بنفس الطريقة السابقة
(اختر مثلا نوع التخطيط : أعمدة Column)
ليظهر لديك شكل مشابه للتالي :



- **اختبار الفرضية** $H_0 : \beta_1 = 0$ **مقابل الفرضية :** $H_1 : \beta_1 \neq 0$ **:**

وتعني الفرضية العدم أنه لا توجد علاقة خطية بين المتغيرين X,Y ، و الفرضية البديلة تعني أنه توجد علاقة خطية.
لإجراء مثل هذا الاختبار سنحتاج إلى جدول تحليل التباين :

$$\sum_{i=1}^{n} y_i = TSS = SSR + SSE = \sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 + \sum_{i=1}^{n} e_i^2 = \beta_1^2 \sum_{i=1}^{n} (X_i - \overline{X})^2 + \sum_{i=1}^{n} e_i^2$$

$$MES = \frac{SSR}{k-1} = \frac{SSR}{1} \quad , \quad MSE = \frac{SSE}{n-2} = S_{YX}^2$$

مؤشر الاختبار : $F = \frac{MRS}{MES}$ ، فإذا كان $F > F_1^{n-2}$ نرفض الفرضية العدم أي هناك علاقة خطية ، و إن كان $F < F_1^{n-2}$ نقبل الفرضية العدم (أي لا توجد علاقة خطية)
جدول تحليل التباين :

S.O.V.	df	SS	MS	F
Regression	1	SS**R**	MSR	$F = \frac{MRS}{MES}$
Residuals (**E**rrors)	n-2	SS**E**	MSE	
Total	n-1	SS**T**		