

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

▪ 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 :

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

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

مؤشر الاختبار : $F = \frac{MRS}{MES}$ ، فإذا كان $F > F_1^{n-2}$ نرفض الفرضية العدم أي هناك علاقة خطية ، وإن كان

$F < F_1^{n-2}$ نقبل الفرضية العدم (أي لا توجد علاقة خطية)
جدول تحليل التباين :

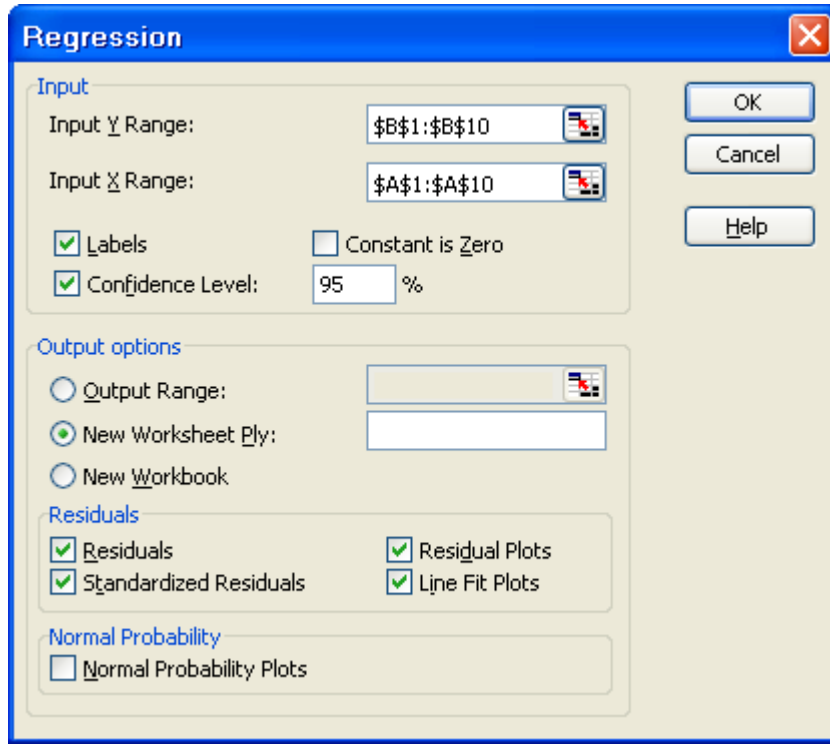
| S.O.V. | df | SS | MS | F |
|--------------------|-----|-----|-----|-----------------------|
| Regression | 1 | SSR | MSR | $F = \frac{MRS}{MES}$ |
| Residuals (Errors) | n-2 | SSE | MSE | |
| Total | n-1 | SST | | |

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

| | 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 | | | | | |

1. لتكن لدينا البيانات الموضحة في الشكل المجاور:

2. من شريط القوائم : أدوات (Tools) < Data Analysis < Regression : فتظهر لدينا النافذة التالية :



في حقل Input Y Range نكتب نطاق خلايا قيم المتحول التابع (Y (Dependent Variable) في حقل Input X Range نكتب نطاق خلايا قيم المتحول المستقل (X (Independent Variable) كما ترى في الشكل فإن الخيار Labels مفعل ، ذلك لأنني بدأت تحديد نطاق الخلايا لكل من x و y ابتداء من عناوين الأعمدة (التي تمثل قيما نصية).

إذا قمنا بتفعيل الخيار Confidence Level فسيتم إظهار مجال الثقة للمعاملات المقدرة بالسنة للرواسب (البواقي) :

بتفعيل الخيار Residuals سيتم إظهار جدول الرواسب، وبتفعيل الخيار Standardized Residuals سيتم إظهار جدول للرواسب المعيارية (سيتم توضيحهم لاحقا)، وبتفعيل الخيار Residual Plots سيتم تمثيل الرواسب بيانيا ، وبتفعيل الخيار Line Fit Plots سيتم رسم لوحة الانتشار للمتحولين X,Y ، إضافة إلى النقاط المقدرة (مستقيم الانحدار). بعد أن تتأكد من المدخلات قم بالنقر على OK ، لتظهر لديك أربعة جداول مذهلة (تجبرك على خلع القبة لشركة مايكروسوفت)

الجدول الأول: Regression Statistics - إحصاءات الانحدار-

SUMMARY OUTPUT

| Regression Statistics | | |
|-----------------------|-------------|-------------------------------------------------|
| Multiple R | 0.965950001 | القيمة المطلقة لمعامل الارتباط |
| R Square | 0.933059405 | معامل التحديد (وهو مربع معامل الارتباط) ... |
| Adjusted R Square | 0.923496463 | يفيد في حالة الارتباط المتعدد |
| Standard Error | 4.702076381 | الخطأ المعياري للتقدير ، أي هو \sqrt{MRS} ... |
| Observations | 9 | 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 $P - value (Sign. F) = 2.31997E - 05 = 2.31997 \times 10^{-5} = 0.0000231997 < \alpha = 0.05$ 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 $\alpha = 0.05$

| | | Coefficients | Standard Error | t Stat | P-value | Lower 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 المقدره وفقا لطريقة المربعات الصغرى | الخطأ المعياري للعامل المقدر | 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 : $H_1 : \beta_0 \neq 0$ وذلك بالنسبة للمعامل $\hat{\beta}_0$ الذي في السطر الثاني .

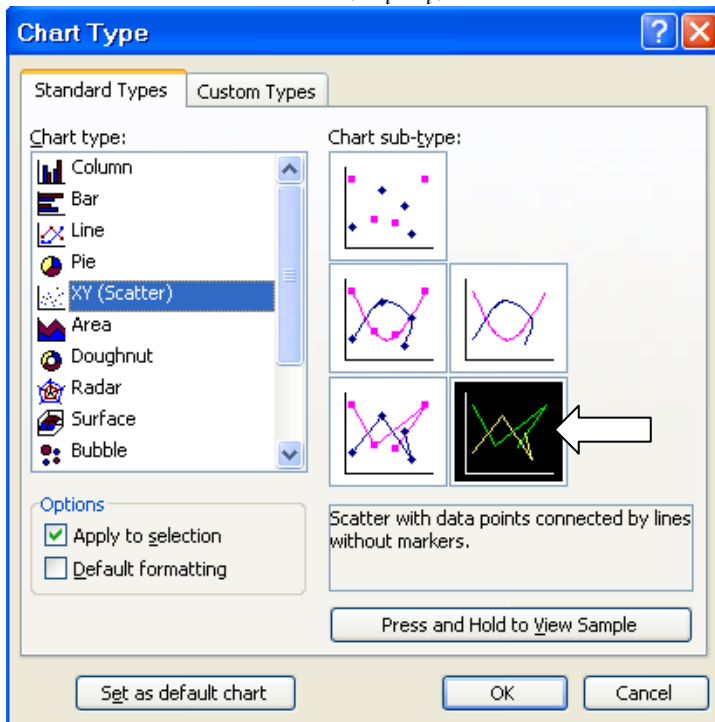
$H_0 : \beta_1 = 0$ مقابل الفرضية البديلة : $H_1 : \beta_1 \neq 0$ وذلك بالنسبة للمعامل $\hat{\beta}_1$ الذي في السطر الأول.

الجدول الرابع : جدول الراسب و التقديرات :

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

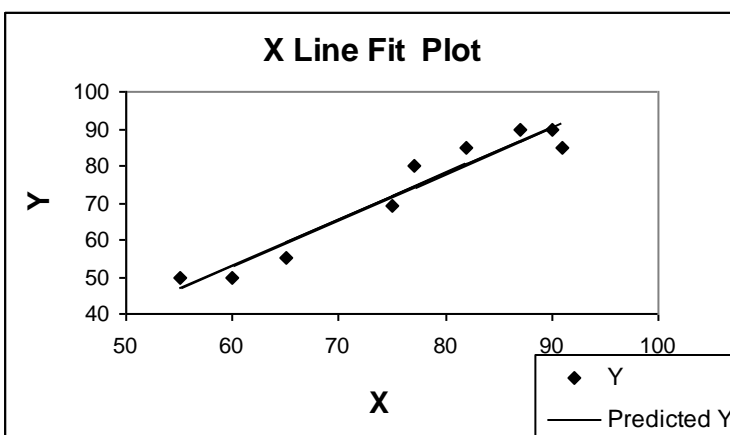
اخيرا يتبقى لدينا الرسوم البيانية :

المخطط البياني الأول هو : X Line Fit Plot ، النقط التي باللون النفسجي تمثل النقط: (X_i, Y_i) ، أما النقط باللون

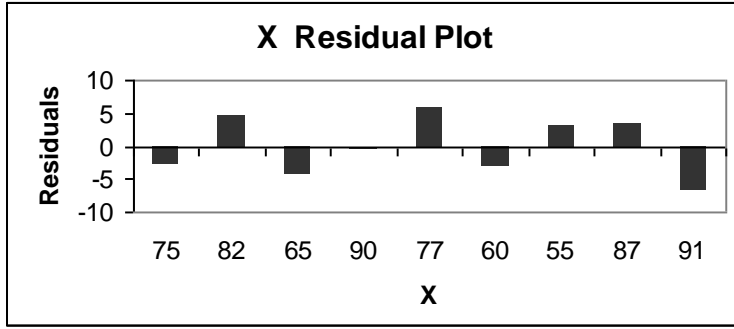


الوردي (Predicted Y) فهي تمثل النقط المقدره وفقا لمعادلة الانحدار ، لاحظ أنها تشكل مستقيما . حتى يكون الشكل مألوفاً أكثر أنصحك بالتالي : انقر بالزر الأيمن للفأر فوق إحدى النقط المقدره (اللون الوردي) عندها تظهر لديك قائمة مختصرة ، اختر منها " نوع التخطيط " "Chart Type" ستظهر لديك النافذة :

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 :

• الانحدار الخطي المتعدد Multiple Linear Regression :

يتم إجراء تحليل الانحدار الخطي المتعدد بنفس الطريقة تقريبا ، والنتائج ستكون مشابهة للسابق ، على سبيل المثال إذا كان لدينا متحول مرتبط Y ، ومعه متحولين مستقلين X1,X2 ، كما هو مبين في الشكل :
من شريط القوائم :

أدوات (Tools) < Data Analysis < Regression

في مربع النص Input Y Range نقوم بتحديد الخلايا من C1 حتى C10 ، وفي مربع النص Input X Range نقوم بتحديد الخلايا من A1 حتى B10 ، كما هو موضح :

| | A | B | C | D |
|----|----|----|-----|---|
| 1 | x1 | x2 | y | |
| 2 | 75 | 69 | 70 | |
| 3 | 82 | 85 | 77 | |
| 4 | 65 | 55 | 80 | |
| 5 | 90 | 90 | 90 | |
| 6 | 77 | 80 | 60 | |
| 7 | 60 | 50 | 110 | |
| 8 | 55 | 50 | 100 | |
| 9 | 87 | 90 | 101 | |
| 10 | 91 | 85 | 112 | |
| 11 | | | | |
| 12 | | | | |

Input

Input Y Range:

Input X Range:

المخرجات ستكون مألوفة لديك .

إن كان لديك أكثر من متحولين مستقلين (ومتحول مرتبط وحيد دائما) يمكنك إجراء تحليل الانحدار المتعدد بنفس الطريقة السابقة ، المهم قم بكتابة بيانات المتحول المستقل في أعمدة متجاورة ، ثم قم بتحديد كلها في الحقل Input X Range

المخططات البيانية Charts

• رسم لوحة الانتشار لمتحولين X,Y Scatter Plot :

لتكن البيانات الموضحة في الشكل المجاور :

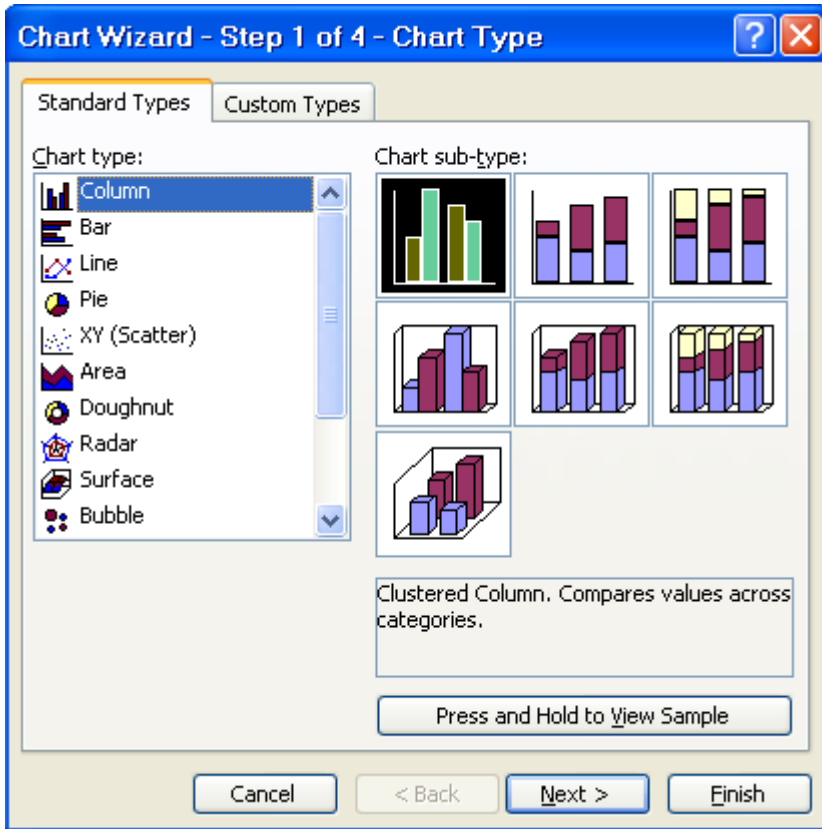
لرسم لوحة الانتشار للنقاط (x,y) نقوم بالتالي :

من شريط القوائم :

إدراج < تخطيط Insert > Chart

ستظهر لدينا النافذة التالية :

| | A | B | C | D |
|----|----|----|---|---|
| 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 | 60 | 60 | | |
| 11 | 91 | 85 | | |
| 12 | | | | |
| 13 | | | | |



لإنجاز الرسم علينا المرور بأربع خطوات :
الخطوة 1 من 4 - نوع التخطيط (Chart Type)
 قم باختيار نوع التخطيط XY (Scatter)
 (س و ص مبعثر)
 انقر على : التالي Next

الخطوة 2 من 4 - البيانات المصدر (Source Data)

في الخطوة الثانية لديك نافذتين
 ميويتين :

نطاق البيانات Data Range و
 متسلسلة Series

انقر على التبويب: متسلسلة Series
 الآن انقر على إضافة (Add) :
 ستظهر لديك ثلاثة حقول :
 ضع المؤشر على الحقل : قيم س
 (X Values) : ثم قم بتحديد نطاق
 الخلايا على صفحة إكسل من الخلية
 A2 حتى الخلية A11.
 امسح محتويات الحقل : قيم ص
 (Y Values) ، ثم قم بتحديد الخلايا
 على صفحة إكسل من الخلية B2
 حتى الخلية B11
 اكتب في الحقل : الاسم Nmae
 (x,y)
 ثم انقر على التالي Next

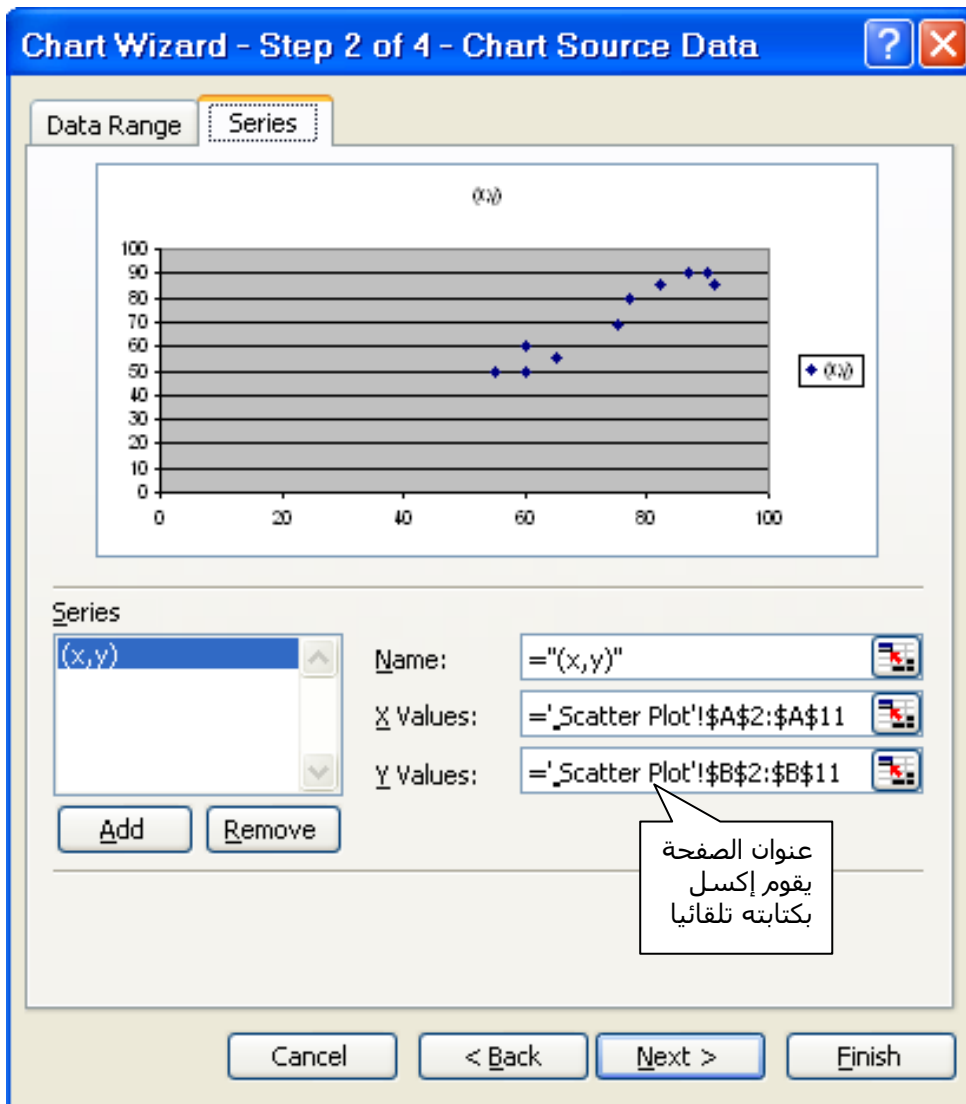
الخطوة 3 من 4 -

خيارات التخطيط Chart Options

يمكنك فيها أن تكتب عنوان التخطيط
 وعناوين المحاور ، انقر على التالي .
 الخطوة الأخيرة هي تحديد مكان إظهار
 التخطيط . اختر المكان ثم انقر إنهاء
 (Finish)

ستظهر عندها لوحة الانتشار في
 ورقة العمل.

يمكنك التعديل على المخطط من حيث:
 اللون ، والأبعاد ، وأحجام النصوص ،
 والترقيم ، وحتى نوع التخطيط ، فقط
 بالنقر مرتين ، أو بالزر الأيمن للفأرة.



| | A | B | C |
|----|----------|-------|---|
| 1 | المبيعات | العام | |
| 2 | 5000 | 1999 | |
| 3 | 6000 | 2000 | |
| 4 | 6500 | 2001 | |
| 5 | 5900 | 2002 | |
| 6 | 7000 | 2003 | |
| 7 | 6700 | 2004 | |
| 8 | 8000 | 2005 | |
| 9 | 9000 | 2006 | |
| 10 | | | |
| 11 | | | |
| 12 | | | |
| 13 | | | |

• التخطيط الشريطي Bar Chart :

لتكن لدينا البيانات الموضحة في الشكل :
لرسم التخطيط الشريطي نقوم باتباع الخطوات :
من: إدراج > تخطيط

نختار في الخطوة الأولى نوع التخطيط : عمودي

في الخطوة الثانية وفي النافذة المبوبة : نطاق البيانات Data Range

وفي الحقل : نطاق البيانات Data Range نقوم بتحديد نطاق الخلايا

للمبيعات (من A2 حتى A9) ، الآن انقر على التبويب "سلسلة Series" في الحقل Name اكتب : " المبيعات " ، و في الحقل :

"عناوين محور س للفئة Category (X) axis labels" قم بتحديد نطاق

الخلايا التي تتضمن الأعوام (أي من B2 حتى B9) ، أكمل باقي

الخطوات كما في السابق لتحصل على الشكل المطلوب.

العمليات على المصفوفات Dealing with Matrixes

سنتناول عمليات الجمع والطرح و الضرب و المنقول و المحدد و المقلوب :

عمليات الجمع و الطرح :

لتكن المصفوفتان الموضحتان في الشكل التالي :

| | A | B | C | D | E | F | G | H | I |
|---|-----|-----|-----|-----|---|-----|----|----|-----|
| 1 | | | | | | | | | |
| 2 | | 9 | 3 | -5 | | | 10 | 11 | 6 |
| 3 | A = | 2 | 4.3 | 0.9 | | B = | 5 | 12 | 16 |
| 4 | | 0.1 | 0 | 5 | | | 4 | -5 | 2.2 |
| 5 | | | | | | | | | |

قبل البدء بعمليات الجمع اتبع الآتي وبدقة :

1- قم بتحديد نطاق عناصر المصفوفة A (في مثالنا من B2 حتى D4)

2- سيظهر لديك في أعلى يسار الشاشة مايلي :

| | A | B | C | D | E | F | G | H | I |
|---|-----|-----|-----|-----|---|-----|----|----|-----|
| 1 | | | | | | | | | |
| 2 | | 9 | 3 | -5 | | | 10 | 11 | 6 |
| 3 | A = | 2 | 4.3 | 0.9 | | B = | 5 | 12 | 16 |
| 4 | | 0.1 | 0 | 5 | | | 4 | -5 | 2.2 |
| 5 | | | | | | | | | |

في الحقل المشار إليه ستجد مكتوباً B2 (أو D4)

3- قم بمسح محتويات هذا الحقل واكتب فيه A (اسم المصفوفة) ثم اضغط على Enter ، أي بالشكل التالي :

| | A | B | C | D | E |
|---|-----|-----|-----|-----|---|
| 1 | | | | | |
| 2 | | 9 | 3 | -5 | |
| 3 | A = | 2 | 4.3 | 0.9 | |
| 4 | | 0.1 | 0 | 5 | |
| 5 | | | | | |

كرر نفس العملية بالنسبة لعناصر المصفوفة B (أي قم بتسمية نطاق الخلايا من G2 حتى I4 بـ B) لإيجاد مجموع المصفوفتين نتبع الخطوات التالية:

- 1- نحدد نطاق خلايا فارغة أبعادها 3*3 (مساوية لأبعاد المصفوفتين السابقتين) ، ولتكن مثلا من B6 حتى D8.
- 2- اكتب أثناء التحديد الصيغة التالية : =A+B
- 3- نضغط على Ctrl و Shift باستمرار الضغط عليهما نضغط على Enter فتظهر المصفوفة الناتجة في الخلايا الفارغة أما بالنسبة لعملية الطرح فيتم بنفس العملية لكن نضع إشارة الطرح بدلا من الجمع .

عمليات جداء المصفوفات :

- 1- قم بتحديد نطاق من الخلايا موافقة لسعة مصفوفة الجداء التي سنتنتج . ولتكن من B10 حتى D12
- 2- اكتب الدالة التالية : =MMULT(A;B)
- 3- نضغط على Ctrl و Shift و باستمرار الضغط عليهما نضغط على Enter فتظهر مصفوفة الجداء A*B .

منقول مصفوفة :

- 1- قم بتحديد نطاق من الخلايا الفارغة ولتكن مثلا من B14 حتى D16 .
- 2- اكتب الدالة التالية : =TRANSPOSE(A)
- 3- نضغط على Ctrl و Shift و باستمرار الضغط عليهما نضغط على Enter فيظهر منقول المصفوفة A في الخلايا الفارغة.

محدد مصفوفة :

محدد المصفوفة عبارة عن قيمة واحدة ، لإيجادها قم بتحديد خلية فارغة ثم اكتب فيها الدالة =MDETERM(A) ثم بالنقر على Enter تظهر لدينا القيمة 165.92 والتي تمثل محدد المصفوفة A.

مقلوب مصفوفة :

يتم بنفس طريقة إيجاد منقول مصفوفة تماما ، غير أننا نكتب الدالة =MINVERSE(A) .
الشكل التالي يوضح العمليات التي قمنا بها :

| | A | B | C | D | E | F | G | H | I | J |
|----|----------|----------|----------|----------|---|----------|----------|----------|----------|---|
| 1 | | | | | | | | | | |
| 2 | | 9 | 3 | -5 | | | 10 | 11 | 6 | |
| 3 | A = | 2 | 4.3 | 0.9 | | B = | 5 | 12 | 16 | |
| 4 | | 0.1 | 0 | 5 | | | 4 | -5 | 2.2 | |
| 5 | | | | | | | | | | |
| 6 | | 19 | 14 | 1 | | | -1 | -8 | -11 | |
| 7 | A+B = | 7 | 16.3 | 16.9 | | A - B = | -3 | -7.7 | -15.1 | |
| 8 | | 4.1 | -5 | 7.2 | | | -3.9 | 5 | 2.8 | |
| 9 | | | | | | | | | | |
| 10 | | 85 | 160 | 91 | | | 112.6 | 77.3 | -10.1 | |
| 11 | A*B = | 45.1 | 69.1 | 82.78 | | B*A = | 70.6 | 66.6 | 65.8 | |
| 12 | | 21 | -23.9 | 11.6 | | | 26.22 | -9.5 | -13.5 | |
| 13 | | | | | | | | | | |
| 14 | | 9 | 2 | 0.1 | | | 10 | 5 | 4 | |
| 15 | A^t = | 3 | 4.3 | 0 | | B^t = | 11 | 12 | -5 | |
| 16 | | -5 | 0.9 | 5 | | | 6 | 16 | 2.2 | |
| 17 | | | | | | | | | | |
| 18 | Det(A) = | 165.92 | | | | Det(B) = | 1209 | | | |
| 19 | | | | | | | | | | |
| 20 | | 0.129581 | -0.09041 | 0.145853 | | | 0.088007 | -0.04483 | 0.086022 | |
| 21 | A^-1 = | -0.05973 | 0.274229 | -0.10909 | | B^-1 = | 0.043838 | -0.00165 | -0.10753 | |
| 22 | | -0.00259 | 0.001808 | 0.197083 | | | -0.06038 | 0.07775 | 0.053763 | |
| 23 | | | | | | | | | | |
| 24 | | | | | | | | | | |
| 25 | | | | | | | | | | |

لمحة عن برنامج Minitab

يعد برنامج Minitab أسهل البرامج الإحصائية الجاهزة على الإطلاق! وهو يشمل كل ما يخطر على بالك من مواضيع الإحصاء التطبيقي. ولكن ما يميزه أيضا هو قسم المساعدة التفاعلي المدعم بهذا البرنامج. من شريط القوائم في برنامج Minitab :

Help > Contents > Menu Commands > Stat

تجد فيه شرحا عن العمليات الإحصائية المتاحة في البرنامج مما ستجد في هذا القسم :

Basic Statistics : الإحصاءات الأساسية (الإحصاء الوصفي ، اختبار الفرضيات)

Regression : الانحدار

ANOVA (Analysis of Variance) : تحليل التباين

DOE (Design of Experiments) : تصميم التجارب

Time Series : السلاسل الزمنية

المهم من كل هذا هو الأمثلة : Help > Contents > Examples : الأمثلة هي أفضل طريقة لتعلم هذا البرنامج

بقي أن أقول لك شيئا لطيفا ! وهو أن المساعدة باللغة الإنكليزية ، ولكن لا عليك فاللغة التي تحتاجها هي لغة بسيطة ، إضافة إلى حصيلة وافرة من المصطلحات العلمية و الإحصائية .

تم بعون الله
معن التنجي- 8- تشرين الأول-2006

المراجع المعتمدة:

1. Excel 2002 لتطبيقات الإحصاء في الأعمال . ترجمة و إعداد هالة الطويل .دار شعاع. حلب 2003.
2. الإحصاء التطبيقي باستخدام Excel . تأليف الدكتور ياسر الموسى ، جامعة حلب . سلسلة الرضا للمعلومات. دمشق 2003.
3. برامج تطبيقية (1) . الدكتور محمد كبيه، الدكتور عبد المجيد بلکش . منشورات جامعة حلب ، مركز التعليم المفتوح 2005.
4. الإحصاء التطبيقي . الدكتور محمد كبيه ، الدكتور ماهر بدوي . منشورات جامعة حلب ، كلية الاقتصاد.
5. قسم المساعدة في برنامج Excel !