شركة **تبارك** للاستشارات ونظم المعلومات
سلسلة التدريب الإحصائي

الإحصاء المهني باستخدام

MS Excel

2003-2007-2010

إعداد: معن التنجي, M.Sc., HND

Tabarac
For Consulting & Information System
Creative Thinking

www.tabarac.com

بسم اللّه الرّحمن الرحيم

"وضرب اللّه مثلاً رجلين أحدهما أبكم لا يقدر على شيءٍ، وهو كلّ على مولاه، أينما يوجّهه لا يأتي بخير، هل يستوي هو ومن يأمر بالعدل وهو على صراط مستقيم"

النحل 76

- الإحصاء المهني باستخدام إكسل
- المؤلف: معن التنجي
- الطبعة الثانية – 2012
- جميع الحقوق محفوظة
- الناشر: شركة تبارك للاستشارات ونظم المعلومات
  - سوريا – حلب
  - هاتف: 1985 577 (21) 963+
  - فاكس: 1986 577 (21) 963+
  - موبايل: 61 63 92 966 963+
  - بريد إلكتروني: info@tabarac.com
  - الموقع على الإنترنت: www.tabarac.com

Tabarac
Consulting & Information System
Creative Thinking

- *Statistics*
- *Business Research*
- *Investment & Islamic Finance*
- *Programming solutions*
- *Information Systems*

تمهيد:

تم تصميم هذا المرجع ليلائم البرنامج التدريبي المتعلق باستخدام الإحصاء في إكسل. غير أنني حاولت تمييز هذا الدليل بعرض سريع للإحصاء بأسلوب مهني مختلف عن الأساليب العلمية المتبعة في الجامعات، بحيث أن المهنيين وغير المختصين يمكنهم فهم أسس الإحصاء إضافة إلى قدرتهم على استخدام برنامج إكسل في مبادئ الإحصاء.

يتميز إكسل بتوفره عموما في أي شركة أو مؤسسة. على خلاف البرامج الإحصائية الجاهزة مثل SPSS والتي قليلا ما تقتنيها الشركات نظرا لكونها برامج تخصصية، فضلا عن ارتفاع سعرها. لذا تبرز هنا أهمية إكسل لكونه متاحا لغالبية الموظفين والمهنيين الذي يستخدمونه في عملهم، فضلا عن سهولة وسرعة تعلمه، الأمر الذي يتيح لهم ايضا استخدامه في حل بعض قضايا بحوث الأعمال التي تتطلب أدوات إحصائية.

غطى هذا المرجع معظم الأدوات الإحصائية الموجودة في إكسل بدءا من الإصدار 2003 وحتى الإصدار 2010 والتي وجدنا أنها الأكثر احتياجا من قبل المهنيين غير المختصين، أو الإحصائيين العاملين.

معن التنجي

حلب – 23 / 4 / 2012

جدول المحتويات:

الإحصاء الوصفي Descriptive Statistics 5

1-1 ما هو الإحصاء الوصفي؟ 5

2.1 الإحصاء الوصفي في برنامج إكسل 2010: 6

1-2-1 مقاييس الإحصاء الوصفي : 7

الجداول التكرارية Frequency Tables : 13

اختبار الفرضيات Test of Hypothesis 15

1.2 ما هي اختبار الفرضيات؟ 15

2.2 طريقة أسهل في اختبار الفرضيات – طريقة الـ P-value: 16

3.2 اختبارات الفرضيات: 16

1.3.2 اختبارات حول متوسط مجتمع – اختبار ت : 16

2.3.2 اختبارات حول الفرق بين متوسطي مجتمعين طبيعيين : 18

3.3.2 اختبار الفرضية حول متوسطي زوج من البيانات لمجتمع ما: 25

تحليل التباين Analysis Of Variance (ANOVA) 27

1.3 تحليل التباين أحادي الاتجاه: 28

1. الإحصاء الوصفي Descriptive Statistics

1-1 ما هو الإحصاء الوصفي؟

لنفرض أنك دخلت إلى مؤسسة ما وسألت أحد مدرائها: هل لك أن تعطيني فكرة عن أعمار العاملين في المؤسسة، فأعطاك قائمة بأعمارهم:

جدول 1

60	55	53	5	30	60	55	45	30	60	45	30	60
33	30	40	53	23	33	20	53	23	33	53	23	33
30	25	25	27	23	30	25	22	23	30	22	23	30
23	28	35	29	30	23	22	19	30	23	19	30	23
30	25	25	30	25	30	25	30	25	30	30	25	30
33	23	40	33	20	33	20	33	20	33	33	20	33
25	25	40	23	34	25	22	23	22	25	23	22	25
49	51	55	33	40	53	70	33	40	53	33	40	53
30	25	25	30	25	30	25	30	25	30	30	25	30
33	25	40	33	20	33	20	33	20	33	33	20	33
25	24	23	23	22	25	22	23	22	25	23	22	25
53	65	60	33	40	53	70	33	40	53	33	40	53
30	25	29	30	25	30	25	30	25	30	30	25	30
33	23	40	33	20	33	20	33	20	33	33	20	33
33	23	40	33	20	33	20	33	20	33	33	20	33
25	23	40	23	22	25	22	23	22	25	23	22	25
53	60	60	33	40	53	35	33	40	53	33	40	53
25	25	25	30	25	30	25	30	25	30	30	25	30
53	65	60	33	40	53	70	33	40	53	33	40	53
30	25	29	30	25	30	25	30	25	30	30	25	30
33	23	40	33	20	33	20	33	20	33	33	20	33
33	23	40	33	20	33	20	33	20	33	33	20	33
25	23	40	23	22	25	22	23	22	25	23	22	25
53	60	60	33	40	53	35	33	40	53	33	40	53

من خلال قراءتك للجدول، هل تستطيع إعطاء أية فكرة عن البيانات، بمعنى هل تستطيع أن تقول مثلا أنهم يميلون للشباب، أم أنهم متقدمون في العمر؟ هذه البيانات تسمى بيانات خام لا يمكن الاستفادة منها إلا من خلال معالجتها. يعتبر الإحصاء الوصفي من أبسط طرق معالجة البيانات للحصول على معلومات مفيدة. فمثلا لو قلت لك بأن متوسط أعمارهم هو 31 سنة، فأنت تحصل هنا على معلومة ذات فائدة "العاملون في المؤسسة يميلون للشباب". لكن هل يكف هذا المقياس لوصف كامل البيانات؟ لنلق نظرة على مجموعتي

البيانات: 1، 30، 60 و 29,30,31 نلاحظ أن لكليهما نفس المتوسط وهو 30، لكن من الواضح أن هناك اختلافا كليا بينهما. لذا عادة ما نستخدم مقياس الانحراف المعياري مع المتوسط لوصف البيانات، فيمكن هنا أن نعبر عن المجموعتين بالمقاييس: 30±30 و 30±1 على الترتيب. فيدل كبر الانحراف المعياري على تشتت أكبر للبيانات عن متوسطها، بينما يدل صغر قيمته على تمركز أكبر للبيانات حول متوسطها.

لكن هل المتوسط هو دائما مقياس يفي بالغرض لوصف مكان تركز البيانات؟

لنلق نظرة على مجموعة بيانات الدخل السنوي بالدولار لعشرة اشخاص تم اختيارهم عشوائيا من إحدى العائلات:

3000، 4000، 4500، 5000، 6000، 6500، 8000، 8500، 9000، 1000000

المتوسط الحسابي (الذي هو مجموع هذه الأرقام مقسوما على عشرة) هو 105450$. من الواضح أن هذا الرقم لا يمثل أبدا بيانات دخل هذه العائلة، السبب ببساطة هو وجود شخص ثري جدا قد أثرت قيمة دخله السنوي على المتوسط الحسابي. الحل هنا هو اعتماد أحد المقياسين، الأول هو المتوسط المشذب Trimmed Mean وهو متوسط الأعداد بعد حذف أكبر 2.5% من البيانات وأصغر 2.5% من البيانات ويسمى هنا المتوسط المشذب – 5% ، أو بالحالة العامة حذف α% من البيانات فيمسى المتوسط المشذب – α%.

أما المقياس الثاني فهو الوسط (أو الوسيط) Median ويحسب من خلال الترتيب التصاعدي للبيانات، فإذا كان عدد البيانات فرديا فالوسط هو القيمة التي في المنتصف تماما، مثال: وسط الأعداد 1، 3، 4، 7، 9 هو 4. أما إذا كان عدد البيانات زوجيا، كبيانات الدخل السابقة، فإن الوسط هو متوسط القيمتين اللتين في المنتصف، أي الوسط هنا هو متوسط القيمتين: 6000 و 6500 وهو 6250$.

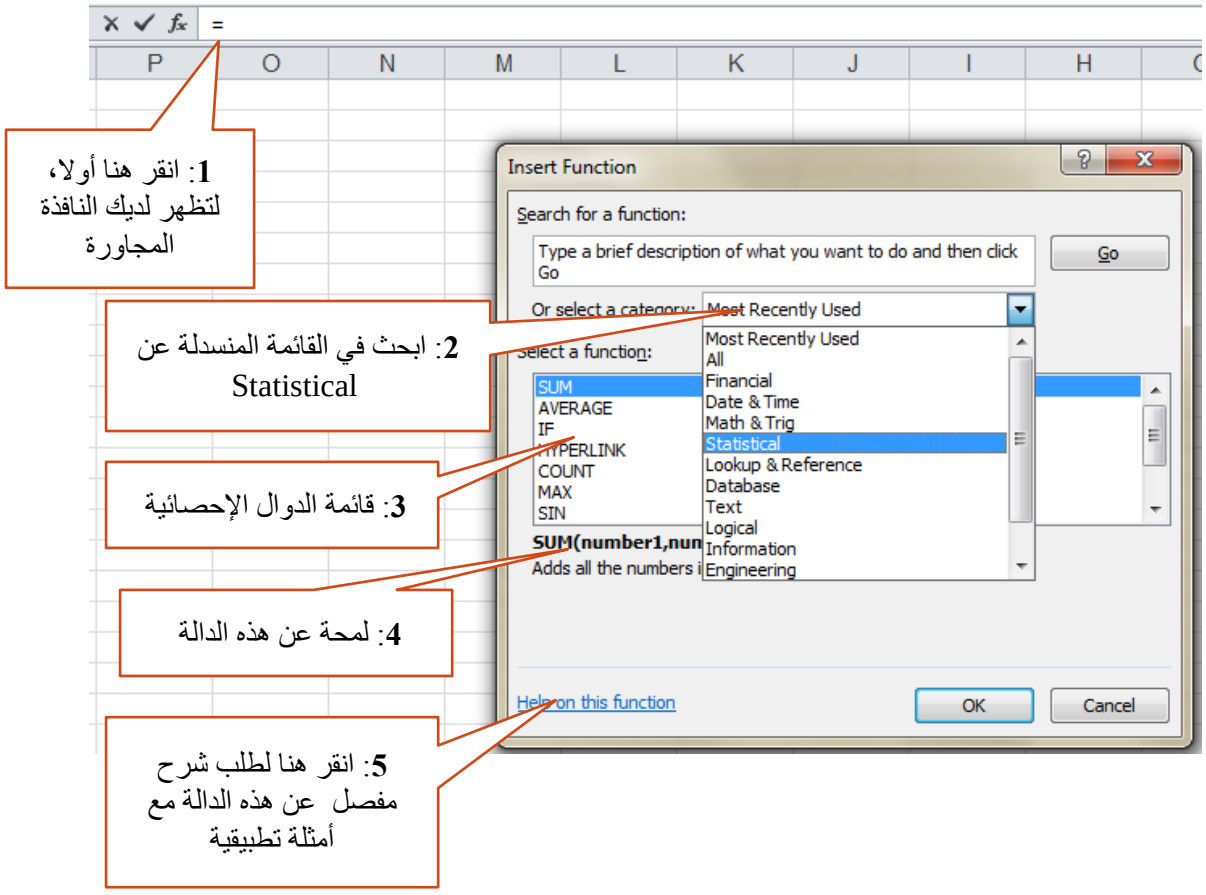
تصنف مقاييس الإحصاء الوصفي في المجموعات الآتية:

- مقاييس النزعة المركزية Central Tendency Measures
- مقاييس التشتت Dispersion Measures
- مقاييس التوزيع Distribution Measures
- الجداول التكرارية Frequency Tables

2.1 الإحصاء الوصفي في برنامج إكسل2007 - 2010:

يمتلك إكسل مجموعة ضخمة من الدوال الإحصائية، سنستعرض حاليا الأكثر استخداما، لكن على العموم تستطيع استعراض كل الدوال الإحصائية الموجودة في إكسل من خلال الخطوات المبينة في الشكل (1) الآتي:

الشكل (1)

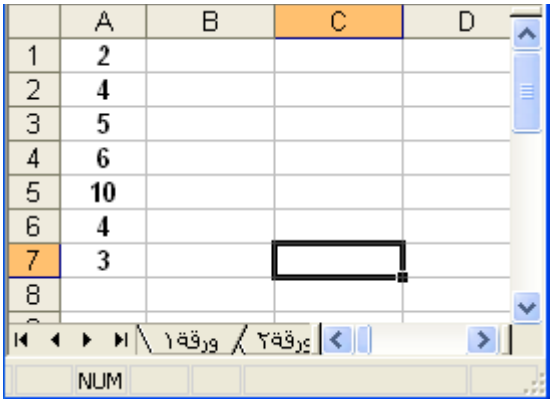


الحقيقة أن قسم المساعدة (الموضح في الخطوة 5) يحتوي شرحا مفصلا وأمثلة باللغة العربية (في حال كانت نسختك معربة) عن كل الدوال، والتي تغني عن أي مرجع. لذا لن أتوسع في شرحها، إنما سأعرض فقط مواطن استخدامها.

1-2-1 مقاييس الإحصاء الوصفي :

لتكن لدينا مجموعة الأعداد المدخلة في صفحة إكسل (الشكل 2):

الشكل 2



لحساب المتوسط لمجموعة الأعداد قم بتحديد أية خلية فارغة ولتكن مثلا C7 ثم اكتب فيها:

=Average(A1:A7)

بالنقر على المفتاح Enter يتم طباعة المتوسط في الخانة C7 وهي القيمة 4.857143 .

المقصود بالصيغة A1:A7 هي مجموعة كل الخلايا المحصورة بين الخليتين A1 و A7 .

للحصول على قيمة المتوسط يمكنك أن تكتب الدالة السابقة كلها يدويا (=Average(A1:A7)) كما يمكنك أن تحدد نطاق الخلايا A1:A7 بشكل أسهل باتباع الخطوات:

اكتب أولا **=Average(** ، ثم قم مباشرة بتحديد الخلايا على صفحة إكسل ، ليتم كتابة نطاق الخلايا تلقائيا ، عندها أغلق القوس ، وانقر على Enter .

إذا أردت حساب الوسط (The Median) اكتب في خلية فارغة : =MEDIAN(A1:A7) بعد النقر على Enter تحصل على القيمة 4 .

وبنفس الطريقة يمكنك حساب بقية المقاييس، الجدول الآتي يبين أهمّ الدالات المستعلمة في برنامج إكسل :

جدول 2

مقاييس النزعة المركزية Central Tendency Measures	المتوسط Mean	=AVERAGE(Range)
	الوسط Median	=MEDIAN(Range)
	المنوال Mode	=MODE(Range)
	المتوسط المشذب Trimmed Mean	=TRIMMEAN(Range; Percent)
مقاييس التشتت Dispersion Measures	التباين Variance	=VAR(Range)
	الانحراف المعياري Standard Deviation	=STDEV(Range)
	الانحراف المتوسط Average of Deviation	=AVEDEV(Range)

	أكبر قيمة Maximum	=MAX(Range)
	أصغر قيمة Minimum	=MIN(Range)
مقاييس التوزيع Distribution Measures	مقياس التفلطح Kurtosis	=KURT()
	مقياس التناظر Skewness	=SKEW()

علما أن إكسل يعتمد الانحراف المعياري المعدل في الدالة STDEV:

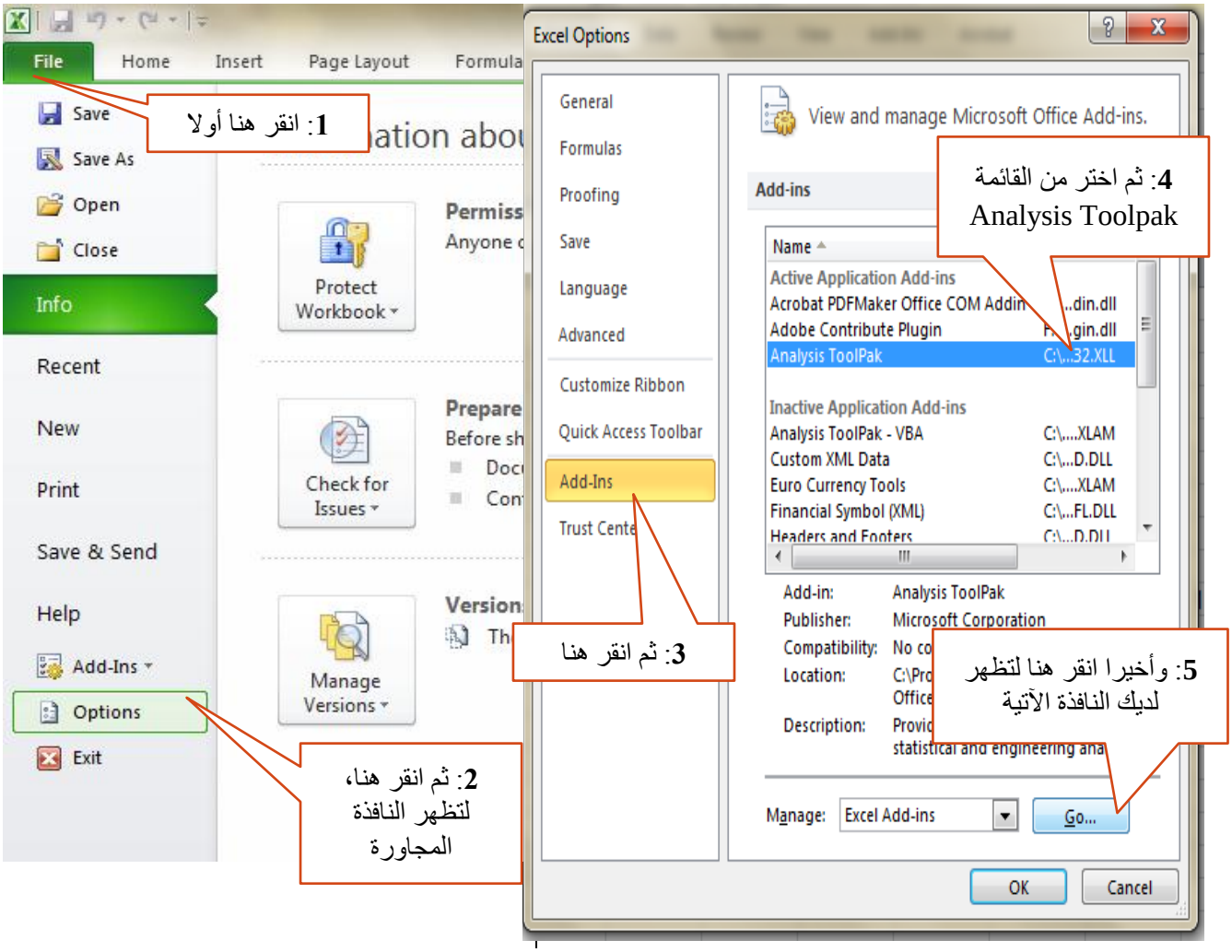
$$SD = \sqrt{\frac{1}{n-1}(x_i - \bar{x})^2}$$

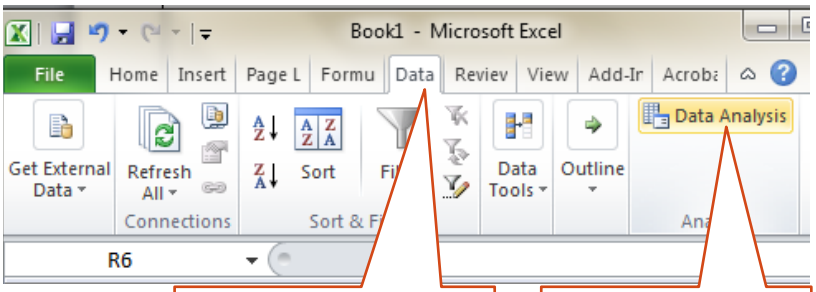
يدلّ كبر قيمة مقياس التفلطح على تفلطح توزيع القيم ، ويدل صغر قيمته على حدة ذروة توزيع القيم.

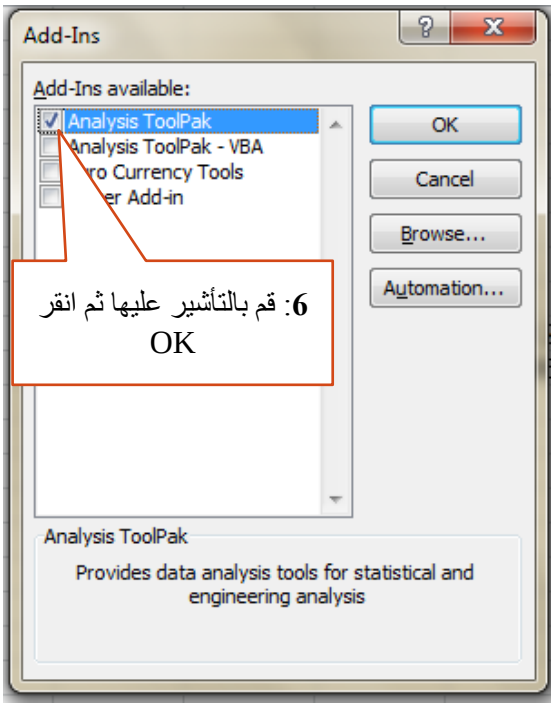
ويدل صغر قيمة عامل التناظر على تناظر التوزيع بالنسبة للتوقع ، حتى إذا وصلت قيمته للصفر دلّ ذلك على تناظر التوزيع بشكل كامل (مثل التوزيع الطبيعي وتوزيع ستيودنت) ، وتدل قيمته الموجبة على انزياح القيم إلى يمين التوقع ، وتدل قيمته السالبة على انزياح القيم إلى يسار التوقع.

يمكنك إظهار كل النتائج السابقة في جدول واحد مرتب ، من خلال وظيفة جاهزة في إكسل.

في البداية ستحتاج من الان فصاعدا إلى تثبيت الوظيفة الإضافية Analysis Tool Pack، الشكل (3) الآتي يوضح طريقة تثبيت الأداة:


File
Home
Insert
Page Layout
Formula
Save
Save As
Open
Close
Info
Recent
New
Print
Save & Send
Help
Add-Ins
Options
Exit
Protect Workbook
Check for Issues
Manage Versions
1: انقر هنا أولا
2: ثم انقر هنا، لتظهر النافذة المجاورة
Excel Options
General
Formulas
Proofing
Save
Language
Advanced
Customize Ribbon
Quick Access Toolbar
Add-Ins
Trust Center
3: ثم انقر هنا
View and manage Microsoft Office Add-ins.
Add-ins
Name
Active Application Add-ins
Acrobat PDFMaker Office COM Addin
Adobe Contribute Plugin
Analysis ToolPak
Inactive Application Add-ins
Analysis ToolPak - VBA
Custom XML Data
Euro Currency Tools
Financial Symbol (XML)
Add-in: Analysis ToolPak
Publisher: Microsoft Corporation
Compatibility:
Location:
Description:
Manage: Excel Add-ins
Go...
OK
Cancel
4: ثم اختر من القائمة Analysis Toolpak
5: وأخيرا انقر هنا لتظهر لديك النافذة الآتية


Book1 - Microsoft Excel
File
Home
Insert
Page L
Formu
Data
Reviev
View
Add-Ir
Acroba
Get External Data
Refresh All
Connections
Sort
Data Tools
Outline
Data Analysis
R6
7: اختر التبويب Data
8: أداة التحليل الإحصائي


Add-Ins
Add-Ins available:
Analysis ToolPak
Analysis ToolPak - VBA
Euro Currency Tools
Add-in
OK
Cancel
Browse...
Automation...
Analysis ToolPak
Provides data analysis tools for statistical and engineering analysis
6: قم بالتأشير عليها ثم انقر OK

أما بالنسبة لإصدار 2003 من الأوفيس فاتبع الخطوات:

من شريط القوائم:

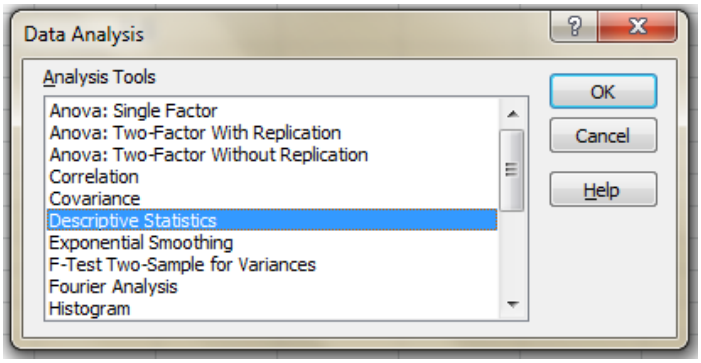
أدوات > وظائف إضافية > Analysis Tool Pack > OK

Tools > Add-Ins > Analysis Tool Pack > OK

قد تظهر لديك رسالة تخبرك أن هذه الوظيفة غير مثبتة لديك، عندها ضع مباشرة القرص الليزري لبرامج الأوفيس، ثم انقر على "موافق" ليتم تثبيتها.

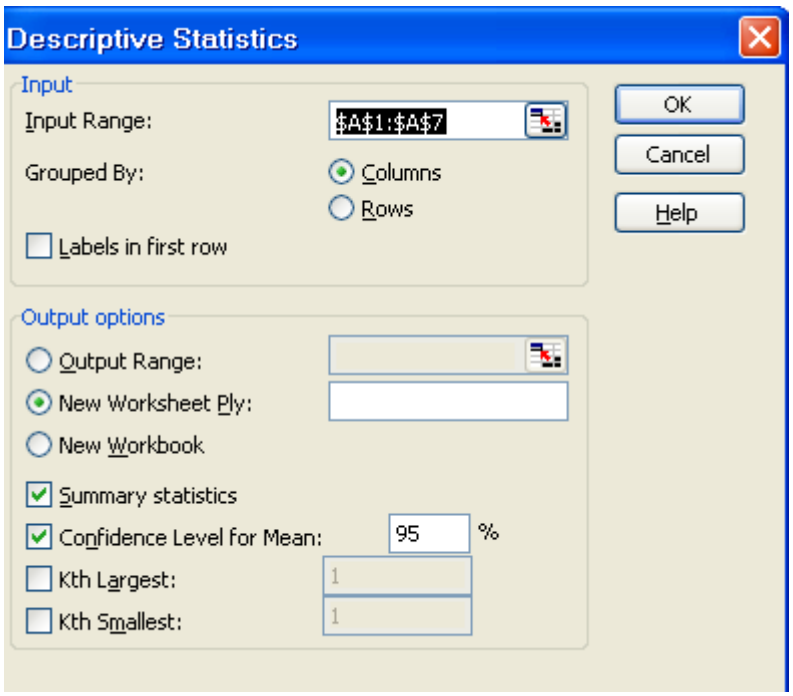
الآن لطلب طباعة جدول الإحصاء الوصفي نفتح نافذة الأداة Data Analysis فتظهر لدينا النافذة الموضحة في الشكل (3) الآتي، والتي سنختار منها Descriptive Statistics:

الشكل (3)



فتظهر لدينا النافذة الآتية:

الشكل 4



ضع المؤشر على مربع النص المقابل لـ Input Range ، ثم على صفحة إكسل قم بتحديد الخلايا من A1 حتى A7 فيتم كتابة نطاق الخلايا بالصيغة المبينة تلقائيا.

قم بتفعيل الخيار Summary Statistics وقم بتفعيل الخيار Confidence Level for Mean ليتم طباعة طول مجال الثقة للمتوسط . بالنقر الآن على OK ستجد جدولا فيه كل الإحصاءات السابقة في صفحة جديدة. كما هو موضح في الشكل 4:

الشكل 5

Column1	
Mean	4.857142857
Standard Error	0.986300723
Median	4
Mode	4
Standard Deviation	2.60950643
Sample Variance	6.80952381
Kurtosis	2.458711917
Skewness	1.398865934
Range	8
Minimum	2
Maximum	10
Sum	34
Count	7
Confidence Level(95.0%)	2.413390927

تفسير المخرجات:

- المتوسط Mean $(\bar{X})$ هو تقريبا 4.86 بانحراف معياري 2.61
- الوسط والمنوال هما 4
- الخطأ المعياري لمتوسط العينة Standard Error هو: 0.99، علما أن الخطأ المعياري لمتوسط العينة له الصيغة: $SE(X) = SD(\bar{X}) = \frac{SD(X)}{\sqrt{n}}$ ، وهو ليس مقياس مكافئا للانحراف المعياري، إنما له استخدامات نظرية إحصائية أخرى (كاستخدامه لحساب مجال الثقة)
- أكبر قيمة هي 10، وأصغر قيمة هي 2، والفرق بينهما (وهو المدى Range) هو 8
- طول مجال 95% ثقة للمتوسط. علما أن الطول يحسب بالصيغة $1.96 \times SE(X)$ علما أن 1.96 هي القيمة الإحصائية الجدولية للتوزيع الطبيعي المعياري المقابلة للاحتمال 95%، وهذه القيمة الثابتة تتغير طبعا وفقا لسوية الثقة التي نريدها. أما مجال الثقة فهو $\bar{X} \pm 1.96 \times SE(X)$، وبالتالي فإن مجال 95% ثقة للمتوسط هو: [2.44, 7.27]. بمعنى أن المتوسط الحقيقي للمجتمع يقع في هذا المجال باحتمال 95%، ويمكننا تجاوزا القول أن 95% من البيانات تقع بين هذين الحدين.

الجداول التكرارية Frequency Tables :

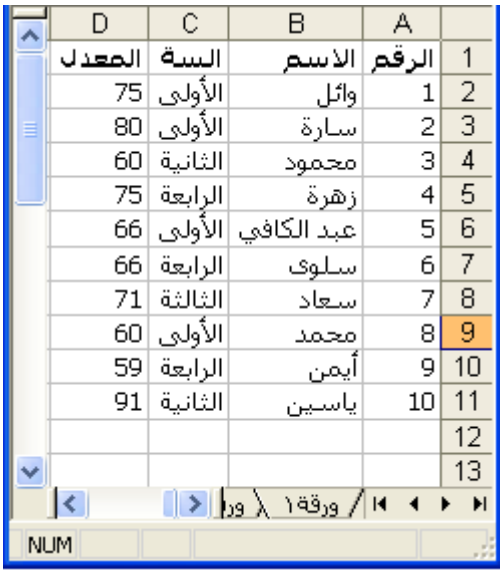
من أجل الجداول التكرارية الوصفية (أحادية القيمة) نستخدم الدالة (Range; cretiria)COUNTIF= لنبين كيفية استخدامها في المثال التاليذ:

لتكن مجموعة الطلاب الموضحة في الشكل 6:

الشكل 6

أ

ب



لإنشاء جدول تكراري بالسنوات الدراسية ، نقوم بالآتي :

أنشئ جدولا في إكسل كالتالي (الشكل 6 – أ)، ثم اكتب في الخلية G8 الدالة التالية : (" الأولى

="؛COUNTIF(C2:C11=، وبنفس الطريقة بالنسبة لباقي السنوات ، حتى تحصل على جدول كالتالي (الشكل 7):

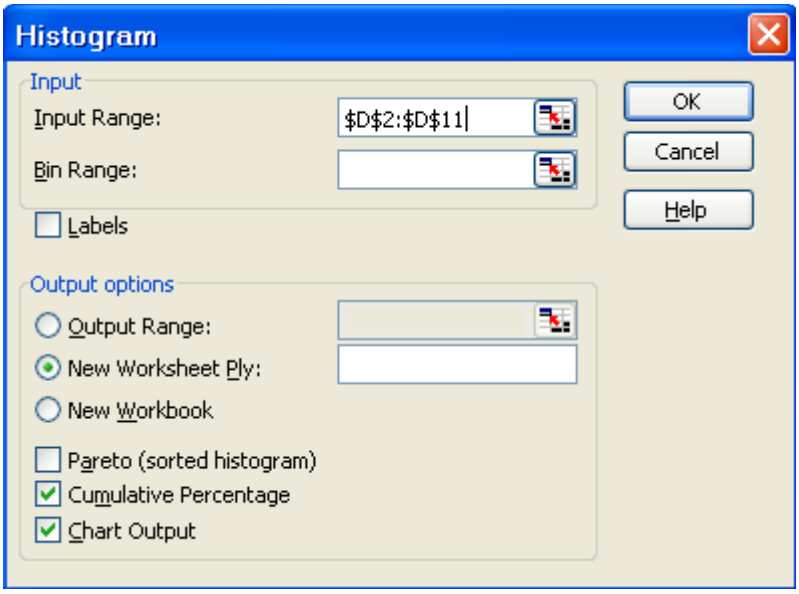
الشكل 7

السنة	عدد الطلاب
الأولى	4
الثانية	2
الثالثة	1
الرابعة	3

الآن لإنشاء جدول تكراري فئوي بالمعدل نقوم بالآتي: Data Analysis < Histogram:

عندها ستظهر لدينا النافذة التالية:

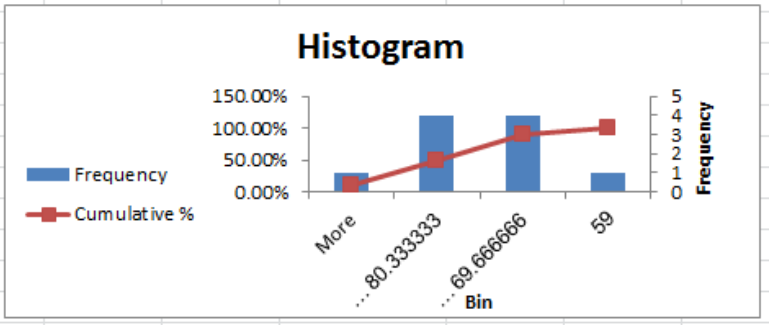
الشكل 8



ضع المؤشر في الحقل Input Range ثم قم بتحديد الخلايا من D2 حتى D11 قم بتفعيل الخيار Cumulative Percentage ليتم طباعة التكرار المئوي المتجمع الصاعد وقم بتفعيل الخيار Chart Output ليتم إظهار التخطيط الشريطي للتكرارات ، انقر على OK ليتم إظهار النتائج في صفحة جديدة. حدود الفئات التي ستظهر قد تم حسابها من قبل إكسل ، إذا أردت تحديد الحدود بنفسك اكتب أطراف مجالات الفئات العليا بشكل عمودي ثم أعد طلب الوظيفة Histogram ثم أعد إدخال البيانات السابقة ، بالإضافة لذلك أدخل في الحقل Bin Range نطاق الخلايا التي تحتوي حدود الفئات عندها سيظهر لديك الجدول التكراري بالفئات التي اخترتها. فيظهر الخرج الآتي (الشكل 9):

الشكل 9

Bin	Frequency	Cumulative %
59	1	10.00%
69.66667	4	50.00%
80.33333	4	90.00%
More	1	100.00%



تفسير المخرجات:

العمود الأول (Bin) هو الحد الأيمن من كل مجال للمعدل، فمثلا القيمة الأولى 59 هي الحد الأيمن من المجال: 59 فأقل، والقيمة 69.7 هي الحد الأيمن من المجال]59 ,69.7].

- العمود الثاني (Frequency) هو عدد البيانات الواقعة في المجال.
- العمود الثالث (Cumulative %) فهو التكرار النسبي المتصاعد.
- أما مخطط الأعمدة فيتناسب طول كل عمود مع التكرارات المقابلة لكل مجال.

2. اختبار الفرضيات Test of Hypothesis

1.2 ما هي اختبار الفرضيات؟

بفرض أننا اطلعنا على أعمار إحدى طلاب السنة الثانية في كلية ما، وكان متوسط أعمار هذه الدفعة الدراسية هو 20.3 سنة، لكن من المفترض أن متوسط أعمارهم يجب أن يكون 20 سنة (أي 20 سنة تمثل عمر الطالب النموذجي الذي لم يرسب أبدا، ودخل السنة الدراسية الأولى وعمره 19 سنة، لا اقل ولا أكثر).

السؤال هنا: هل سنعتبر أن أعمارهم تعتبر تقريبا نموذجية؟ أي هل سنعتبر أن الفرق بين العمر النموذجي الذي ينبغي أن يكون عليه الطلاب هو 20 والعمر الفعلي لهم وهو 20.3 ليس جوهريا. بمعنى هل نستطيع تجاهل الفرق والذي هو 0.3 سنة؟

هذا السؤال يمكن صياغته رياضيا بالشكل:

$$\begin{aligned} H_0&: \bar{X} = 20 \\ H_A&: \bar{X} \neq 20 \end{aligned} \quad \ldots(1)$$

نسمي الفرضية H_0 بالفرضية الابتدائية (العدم)، ونسمي الفرضية H_A بالفرضية البديلة. فإذا أجرينا الاختبار الإحصائي المناسب وقبلنا على أساسها الفرضية الابتدائية نتج أن أعمارهم نموذجية، أي ليس هناك خلل ملحوظ في اعمارهم. أما إذا رفضنا الفرضية الابتدائية فهذا يعني أن أعمارهم ليست نموذجية، بمعنى أن هناك خللا أو مشكلة ما في أعمارهم.

لنوضح أكثر دور اختبار الفرضيات من خلال مثال سبق ذكره عن دخول إحدى العائلات:

3000، 4000، 4500، 5000، 6000، 6500، 8000، 8500، 9000، 100000

الفرضية الابتدائية ستكون هل ينتمي الدخل المميز 100000 إلى مجموعة دخول بقية أفراد العائلة؟ ستكون عندها الفرضية على الشكل:

$$\begin{aligned} H_0&: \bar{X} = 100000 \\ H_A&: \bar{X} \neq 100000 \end{aligned} \quad \ldots(2)$$

فإذا أجرينا أيضا الاختبار الإحصائي المناسب وقبلنا على أساسها الفرضية الابتدائية نتج أن هذا الدخل يقترب من بقية ذخول العائلة، أما إذا رفضنا الفرضية الابتدائية فهذا يعني أن هذا الدخل شاذ، ولا يمثل دخول هذه العائلة، بمعنى أنه عند دراسة دخول هذه العائلة ينبغي استثناء هذا الدخل الشاذ.

2.2 طريقة أسهل في اختبار الفرضيات – طريقة الـ P-value:

لاختبار فرضية عدم مقابل فرضية بديلة يمكننا اتباع إحدى الطريقتين :

الطريقة الأولى: يتم فيها حساب القيمة الفعلية (مؤشر الاختبار) و القيمة النظرية (أي القيمة التي نحصل عليها من الجداول الإحصائية) ، و من خلال مقارنة هاتين القيمتين نتخذ القرار برفض أو قبول الفرضية العدم.

الطريقة الثانية: هي طريقة الـ P-value وهي الطريقة المتبعة في معظم البرامج الإحصائية و ذلك لسهولتها، و يستفاد من هذه القيمة وفق الشكل التالي:

نقارن قيمة P-value مع مستوى المعنوية α ونتخذ القرار وفق الجدول:

جدول 2

$P\text{-value} > \alpha$	نقبل الفرضية العدم
$P\text{-value} < \alpha$	نرفض الفرضية العدم

تختلف قيمة P-value وفقا لاتجاه الفرضية البديلة (غير أنها لا تتعلق أبدا بمستوى المعنوية α) علما أن مستوى المعنوية α عادة ما يؤخذ 0.05، وأحيانا أخرى 0.01

فما هي الـ P-value ؟

احتمال الحصول على قيمة أكبر أو تساوي من (أقل أو تساوي من) إحصائية الاختبار المحسوبة من بيانات العينة آخذاً في الاعتبار توزيع إحصائية الاختبار بافتراض صحة فرض العدم وطبيعة الفرض البديل. على كل حال لست مضطرا لفهم هذه القيمة للاستفادة منها.

3.2 اختبارات الفرضيات:

تختلف طريقة اختبار الفرضية وفقا للحالة المدروسة، وعموما تصنف اختبار الفرضيات إلى صنفين:

1- مجتمع واحد نريد اختبار فيما إذا كان متوسطه يساوي قيمة محددة أم لا (كالفرضية 1 و2).

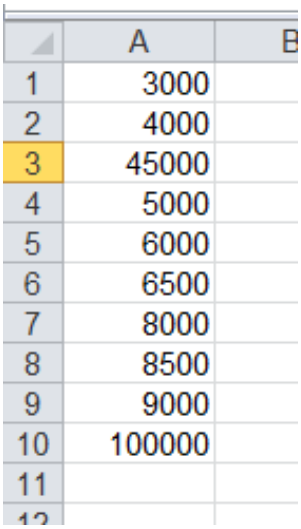
2- مجتمعين ونريد مقارنة متوسطي المجتمعين (أو التباينين أيضا)، وتحديد فيما إذا كان هنالك فرق جوهري بين المجتمعين أم لا.

1.3.2 اختبارات حول متوسط مجتمع – اختبار ت :

لنعد ثانية إلى مجموعة أرقام الدخل ولنفرض أننا أدخلناها في إكسل كما يوضح الشكل (10-أ) الآتي:

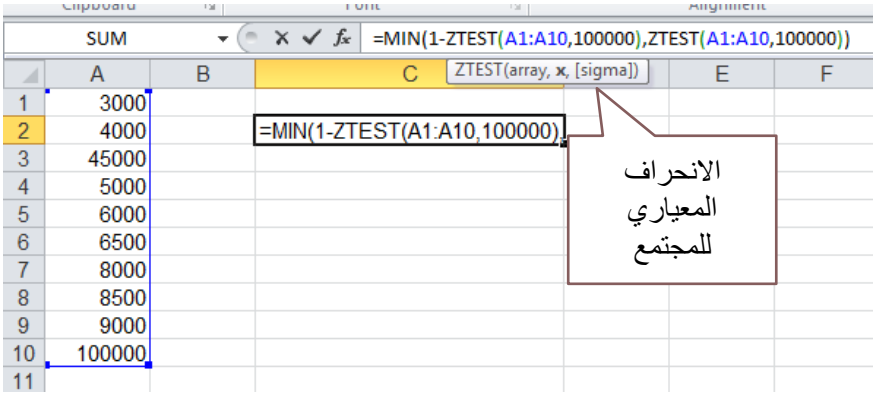
(الشكل 10):

أ



	A
1	3000
2	4000
3	45000
4	5000
5	6000
6	6500
7	8000
8	8500
9	9000
10	100000
11	

ب



بفرض أننا سنختبر الفرضية:

$$H_0: \bar{X} = 100000$$
$$H_A: \bar{X} \neq 100000 \quad \ldots(3)$$

عندها نحن أمام حالة فرضية بديلة ثنائية الذيل، في هذه الحالة اكتب في خلية مجاورة الدالة:

=MIN(1-ZTEST(A1:A10,100000),ZTEST(A1:A10,100000))

ثم انقر على Enter لتظهر لديك القيمة 1.11022E-16، علما أن هذه الصيغة يقصد بها القيمة 1.1×10^{-16} وهي مانسميه قيمة الـ P-value , والتي تعني رفض الفرضية الابتدائية (انظر الجدول 2) القائلة بأن قيمة الدخل 100000 تنتمي لمجتمع دخول العائلة. أي أن هذا الدخل مختلف تماما ولا ينتمي إلى هذا المجتمع.

يشير الوسيط الذي يكتب بين قوسين متوسطين [sigma] والذي هو الانحراف المعياري للمجتمع إلى أن المستخدم مخير في ملء هذا الوسيط (إن كان يعرف مسبقا الانحراف المعياري للمجتمع)، أو يتركه دون ملء، أي يملأ عندها الدالة بوسيطين فقط، فيستبدل إكسل الانحراف المعياري للمجتمع بانحراف العينة.

أشكال أخرى من الفرضية:

تعتبر شكل الفرضية (3) هي أكثر أشكال الفرضيات استخداما، لكن بملاحظة أن الفرضية الابدائية ترفض في حالتين: الأولى عندما تكون 100000 أكبر بكثير من ، والثانية عندما تكون 100000 أصغر بكثير من $\bar{X}$ ، وتحقق الحالية الثانية أمر مستحيل، لذا نحن أمام فرضية بديلة غير مناسبة للحالة المدروسة، بالطبع استخدام الفرضية 3 يعتبر صحيحا ويعطي نتائج صحيحة، لكنها مجملا لا تعبر دقيقة، والأدق هو استخدام الفرضية:

$$H_0: \bar{X} = 100000 \quad \ldots(3)$$

$$H_A: \bar{X} < 100000$$

أي أننا في الفرضية البديلة نفترض أن الدخل 100000 أكبر بكثير من المتوسط $\bar{X}$ ، في هذه الحالة لا تختلف لدينا إلا طريقة حساب المعنوية P-value ، والتي تحسب عموما وفق الجدول:

جدول 3: طرق حساب المعنوية وفقا لشكل الفرضية

الفرضية البديلة	***P-value***
$H_A: \bar{X} = x_0$	=2 * MIN(ZTEST(..) ; 1-ZTEST(..))
$H_A: \bar{X} < x_0$	=1 - ZTEST(..)
$H_A: \bar{X} > x_0$	=ZTEST(..)

إذا، نستخدم الفرضية ثنائية الذيل في الحالات العامة عندما لا نعلم شكل العلاقة بين المتوسط والقيمة المختبرة (أي لا نعلم هل هي علاقة أصغر أو أكبر)، لكن عندما نعلم شكل العلاقة، ونريد اختبار هذه العلاقة، فمن الأفضل استخدام إحدى أشكال الفرضية البديلة لنحصل على نتائج أدق.

2.3.2 اختبارات حول الفرق بين متوسطي مجتمعين:

تختبر هذه الفرضية السؤال: هل هناك فرق معنوي بين متوسطي مجتمعين؟ بحيث تكون الفرضية الابتدائية هي (لا يوجد فرق معنوي بين المتوسطين) مقابل الفرضية البديلة (يوجد فرق معنوي بين المتوسطين، أو أحد المتوسطين أكبر معنويا من الآخر).

لنأخذ التطبيق الآتي:

Are American women's feet getting bigger?

سؤال أصاب حيرة تجار التجزئة في الولايات المتحدة الأمريكية في العشرين سنة الأخيرة. فمثلا مخازن شركة Wal-Mart ومخازن شركة Payless تنفد منها الأحذية ذات المقاس الكبير أسرع من غيرها، مما يطرها لزيادة مخزونها سنويا من الأحذية الكبيرة. فأجريت دراسة للإجابة على هذا السؤال، فأذخت عينة من 12 امرأة من مواليد عام 1980، و 12 امرأة أخرى من مواليد 1960 وأخذت مقاسات أحذيتهم، فكانت البيانات:

جدول 4

مواليد 1980	9	8.5	8	8	7.5	9.5	7.5	8	8.5	8	8.5	8
مواليد 1960	8	7	8	7	7.5	7.5	7.5	8	8	7.5	8	8.5

مصدر الدراسة: Wall Street Journal, July 2004

هل هناك زيادة فعلية؟

يمكننا صياغة هذا السؤال بهذه الفرضية:

- الفرضية الابتدائية: مقاسات أحذية نساء الجيل السابق (مواليد 1980) هي نفسها تقريبا مقاسات أحذية نساء الجيل الحديث (مواليد 1960)
- الفرضية البديلة: مقاسات أحذية نساء مواليد 1980 أكبر من مقاسات أحذية نساء مواليد 1960 (الفرضية البيلة تسمى أحادية الذيل، كما يمكن صياغة الفرضية البديلة كثنائية الذيل بالشكل: مقاسات أحذية نساء مواليد 1980 لا تساوي مقاسات أحذية نساء مواليد 1960، لكن تبقى الفرضية أحادية الذيل أدق في هذه الحالة المدروسة).

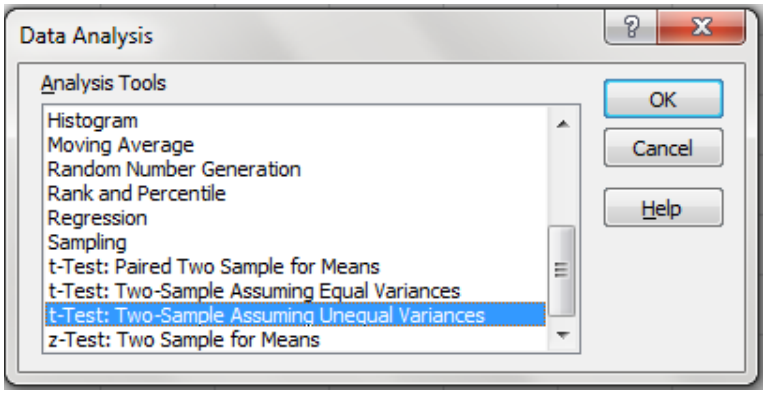
يوضح الشكل الشكل (11) البيانات المدخلة:

الشكل (11)

	A	B	C
1	مواليد 1980	مواليد 1960	
2	8	8.5	
3	7.5	7.5	
4	8.5	8	
5	8.5	8	
6	8	7.5	
7	7.5	7.5	
8	9.5	7.5	
9	7.5	8	
10	8	7	
11	8	8	
12	8.5	7	
13	9	8	
14			
15			

الآن من الأداة Data Analysis فتظهر لدينا النافذة (الشكل 12):

الشكل 12



الاختبار الذي سنستخدمه هو اختبار t-Test ، لكن نجد أمامنا خيارين:

- T-Test: Two-Sample Assuming Equal Variances
- T-Test: Two-Sample Assuming Unequal Variances

كما هو واضح، طريقة الاختبار الأولى تفترض أن تبايني المجتمعين متساويين، والطريقة الثانية تفترض أن التباينين غير متساويين. بحسابنا لقيم الانحراف المعياري لكل جيل من النساء نجد:

مواليد 1960	0.450168
مواليد 1980	0.620056

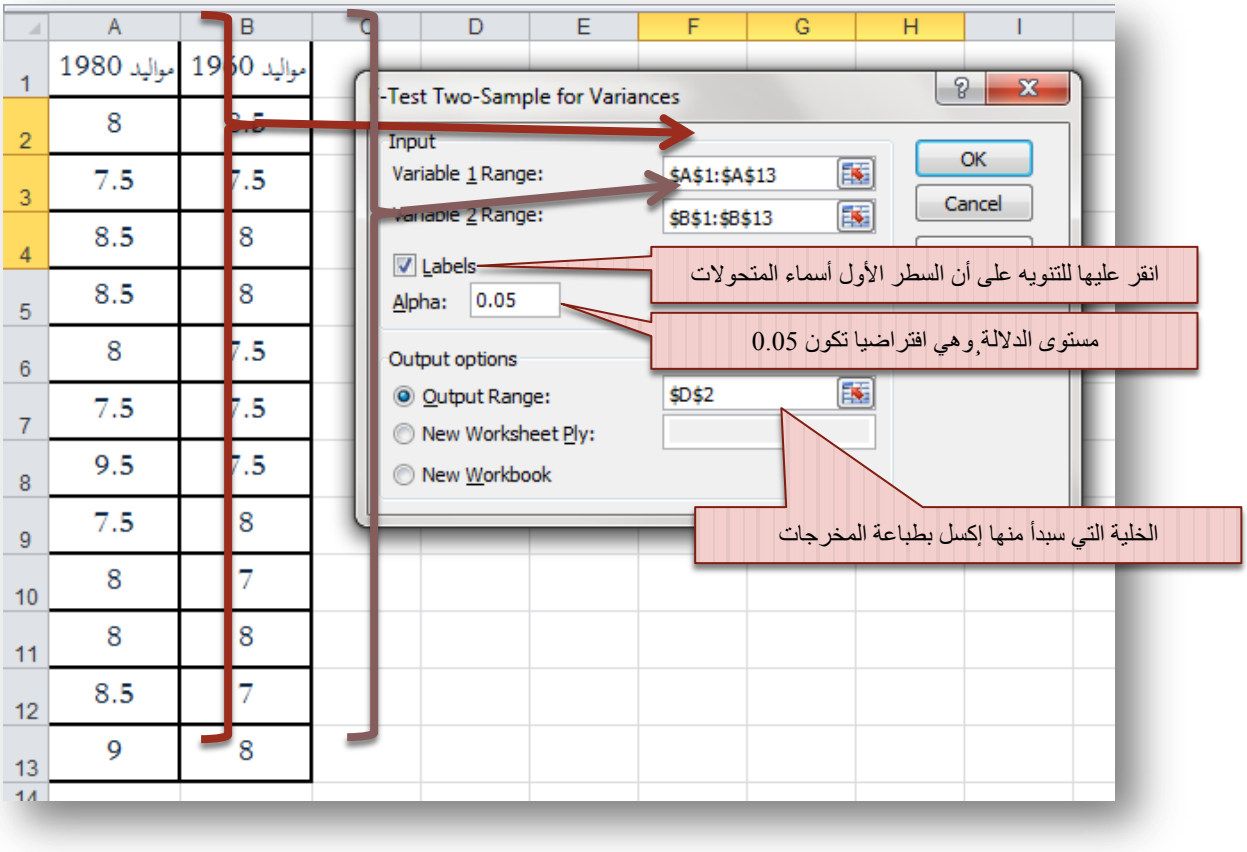
فهل نعتبر أن الفرق بين الانحرافين المعياريين مهمل ونأخذ حالة الاختبار الأول التي تفترض تساوي التباينين، أم نأخذ الاختبار الثاني؟ في حال وضوح التساوي (أو وضوح الفرق الكبير) بين الانحرافين المعياريين، لا مانع من الاختيار المباشر لطريقة الاختبار. لكن عند عدم وضح شكل العلاقة بين الانحرافين، لا بد من إجراء اختبار إحصائي خاص (وهو اختبار F-test Two-Sample for Variances) ويختبر هذا الاختبار الفرضية الآتية:

- **الفرضية الابتدائية**: الانحراف المعياري للمجتمع الأول يساوي الانحراف المعياري للمجتمع الثاني
- **الفرضية البديلة**: الانحراف المعياري للمجتمع أكبر (أو أصغر) من الانحراف المعياري للمجتمع الثاني

أي عندما تكون معنوية الاختبار أكبر من مستوى الدلالة (وليكن مثلا 0.05) عندها نقبل بالفرضية الابتدائية القائلة بتساوي الانحرافين، وعندما تكون المعنوية أصغر من مستوى الدلالة عندها نرفض بالفرضية الابتدائية القائلة وبالتالي لا بتساوى الانحرافان.

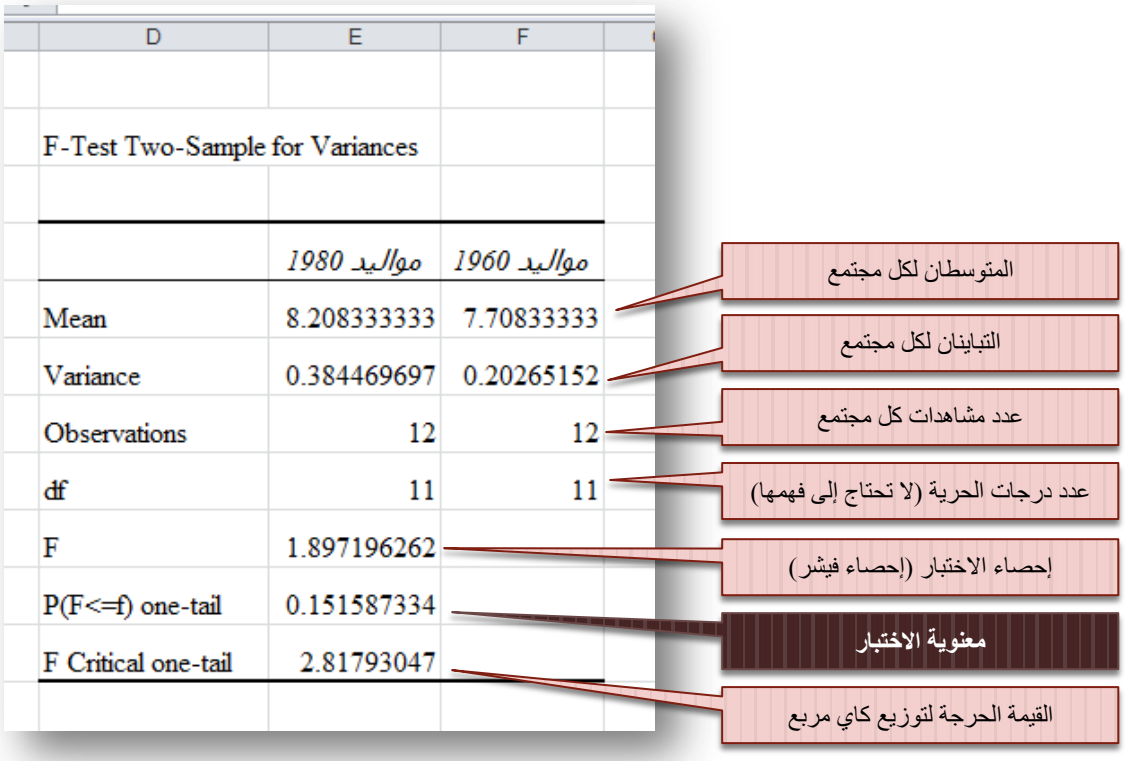
لإجراء هذا الاختبار نختار من الأداة Data Analysis الاختبار F-test Two-Sample for Variances فتظهر لدينا النافذة الآتية (الشكل 13):

الشكل 13


-Test Two-Sample for Variances
Input
Variable 1 Range:
$A$1:$A$13
Variable 2 Range:
$B$1:$B$13
OK
Cancel
Labels
Alpha:
0.05
Output options
Output Range:
$D$2
New Worksheet Ply:
New Workbook
انقر عليها للتنويه على أن السطر الأول أسماء المتحولات
مستوى الدلالة,وهي افتراضيا تكون 0.05
الخلية التي سبدأ منها إكسل بطباعة المخرجات

بالنقر على تظهر لدينا المخرجات الآتية (الشكل 14):

الشكل 14



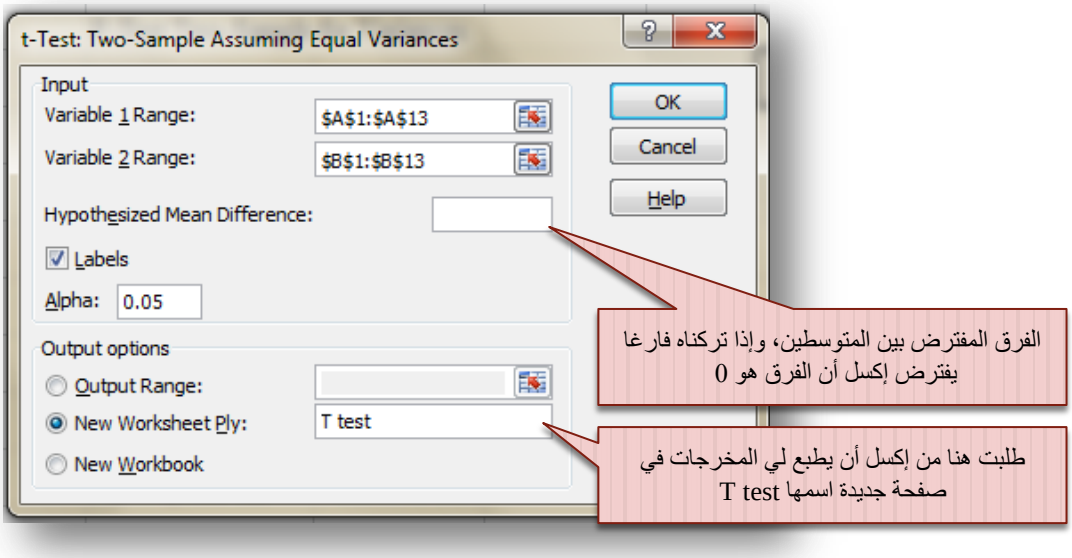
يتضح من معنوية الاختبار (P-value) والتي هي أكبر من مستوى الدلالة الذي اخترناه (0.05) وبالتالي فإننا نقبل بالفرضية الابتدائية القائلة بتساوي تبايني المجتمعين (أو الانحرافين المعياريين).

لنعد الآن إلى تطبيق تغير مقاسات الأحذية، نذكر اننا وقفنا عندها عند خيارين (الشكل 12):

- T-Test: Two-Sample Assuming Equal Variances
- T-Test: Two-Sample Assuming Unequal Variances

من الواضح أننا سنعتمد على الطريقة الأولى التي تفترض تساوي تبايني المجتمعين. لذا باختيارنا لهذه الوظيفية (T-Test: Two-Sample Assuming Equal Variances) في الأداة Data Analysis تظهر لدينا النافذة (الشكل 15):

الشكل 15



تظهر النتائج كالآتي (الشكل 16):

الشكل (16)

	A	B	C
1	t-Test: Two-Sample Assuming Equal Variances		
2			
3		مواليد 1980	مواليد 1960
4	Mean	8.208333	7.708333
5	Variance	0.38447	0.202652
6	Observations	12	
7	Pooled Variance	0.293561	
8	Hypothesized Mean Difference	0	
9	df	22	
10	t Stat	2.26046	
11	P(T<=t) one-tail	0.017014	
12	t Critical one-tail	1.717144	
13	P(T<=t) two-tail	0.034029	
14	t Critical two-tail	2.073873	

تباين الفرق بين المتوسطين ، أي هو: $Var(\bar{X}_1 - \bar{X}_2)$

الفرق المختبر

درجة حرية إحصاء الاختبار

إحصاء الاختبار

المعنوية في حالة الفرضية البديلة أحادية الذيل

القيمة الحرجة في حالة الفرضية البديلة أحادية الذيل

المعنوية في حالة الفرضية البديلة ثنائية الذيل

القيمة الحرجة في حالة الفرضية البديلة ثنائية الذيل

من الواضح أن متوسط مقاسات أحذية نساء مواليد 1980 أكبر من مقاسات أحذية نساء مواليد 1960، لكن هل هذا (الكبر) معنوي، بالنظر إلى قيمة المعنوية P(T<=t) one-tail والتي تساوي 0.017 نجد أن المعنوية أصغر من مستوى الدلالة 0.05،

وبالتالي فإننا نرفض الفرضية الابتدائية القائلة بتساوي مقاسات أحذية الجيلين، ونقبل بذلك الفرضية البديلة القائلة بأن مقاسات الأحذية قد كبرت فعلا بشكل ملحوظ.

وبالطبع يتم تنفيذ وتفسير مخرجات الاختبار T-Test: Two-Sample Assuming Unequal Variances بنفس الطريقة، والحقيقة أن نتائج الاختبارين متقاربة نسبيا، غير أن اختيار الاختبار الأنسب ضرب من الدقة.

الحالة العامة لشكل الفرضية:

اعتبرنا في المثال السابق أن الفرق المفترض بين المتوسطين هو 0، لكن في الحالة العامة يمكن أن نختبر الفرضية:

$$H_0: \bar{X}_1 - \bar{X}_2 = x_0$$
$$H_A: \bar{X}_1 - \bar{X}_2 \quad \neq, <, > \quad x_0 \qquad \ldots(4)$$

أي أننا سنختبر هل الفرق بين المتوسطين هو تقريبا x_0، أم أن الفرق هو مختلف معنويا عن القيمة x_0 ، أي أن الفرق بين المتوسطين يحقق إحدى العلاقات المنطقية: $\neq, <, >$ عند مستوى دلالة. على سبيل المثال، إذا أردنا دراسة الفرق بين أعمار طلاب السنة الثالثة مع طلاب الرابعة، فإننا هنا نختبر: هل الفرق بينهما هو فقط سنة واحدة (أي الفرق بين الدفعتين طبيعي)، أم أن الفرق مختلف معنويا عن سنة (هناك خلل في أعمار غحدى الدفعتين). في مثل هذه الحالة فإننا نملء الخانة Hypothesized Mean Difference عند الإدخال (الشكل 15) بالقيمة المفترضة للفرق بين المتوسطين وهو هنا مثلا (1).

من الملاحظ في المثال السابق أن المجموعتين المدروستين (مواليد 1980 ومواليد 1960) مختلفتان عن بعضهما، بمعنى آخر مستقلتان. تصادفنا حالة أخرى نقارن بها مجموعتين من البيانات لمجموعة واحدة مدروسة في ظرفين مختلفين. على سبيل المثال: عندما نقارن أداء فريق المبيعات قبل الدورة التدريبية وبعدها. نلاحظ أننا هنا أمام مجموعة مدروسة واحدة وهي فريق المبيعات، لكن في زمنين مختلفين (قبل الدورة وبعد الدورة). كمثال آخر، قررت شركة ما تطبيق نظام مراقبة وخفض التكاليف في الشركة (نظام ABC على سبيل المثال)، وأرادت الشركة مقارنة تكاليف الشركة قبل تطبيق النظام وبعد تطبيقه. في مثل هذه الحالة نستخدم اختبار Paired Two-Sample for Means .

3.3.2 اختبار الفرضية حول متوسطي زوج من البيانات لمجتمع ما:

لنوضح هذا الاختبار من خلال التطبيق الآتي:

أرادت شركة أدوية تجريب عقار لتخفيض وزن النساء، فقامت بتجريبه على عينة من النساء حجمها 12، وتم تسجيل أوزانهم، ثم بدأت المجموعة بتناول العقار لمدة شهر، عندها قامت الشركة بتسجيل أوزانهم مرة ثانية لتحديد إن كان هناك تأثير جوهري على تخفيض الوزن أم لا، فكانت الأوزان كالتالي:

قبل استعمال العقار	71	69	92	83	81	77	101	96	88	86	90	80
بعد شهر من استعمال العقار	69	65	89	79	79	76	97	90	85	84	87	75

هل هناك تأثير معنوي للعقار على تخفيض وزن النساء؟

لاحظ من هذا التطبيق، أننا ندرس نفس المجتمع، لكن في ظرفين مختلفين، قبل استعمال العقار، وبعد استعماله. أي أننا سنختبر هنا الفرضية الابتدائية: لا يوجد فرق معنوي بين أوزان النساء قبل وبعد استخدام العقار، مقابل الفرضية البديلة (أحادية الذيل): أوزان النساء بعد استخدام العقار أقل من أوزانهن قبل استخدامه.

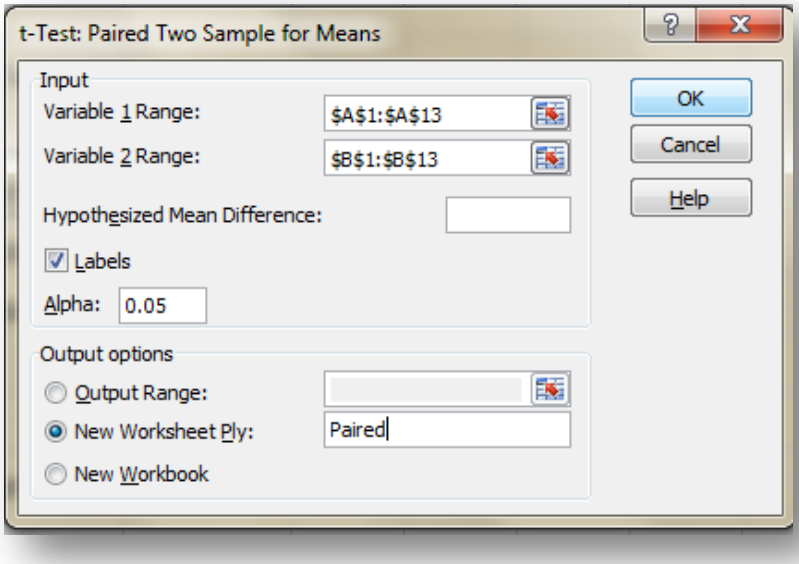
لإجراء هذا الاختبار ندخل البيانات كما هو موضح في الشكل 17 الآتي:

الشكل 17

	A	B
1	قبل استعمال العقار	بعد شهر من استعمال العقار
2	71	69
3	69	65
4	92	89
5	83	79
6	81	79
7	77	76
8	101	97
9	96	90
10	88	85
11	86	84
12	90	87
13	80	75
14		

ثم نختار الوظيفة Paired Two-Sample for Means من الأداة Data Analysis فتظهر لدينا النافذة الآتية (الشكل 18):

الشكل 18



نقوم بملأ المدخلات كما هو موضح (وكما فعلنا في أمثلة سابقة) فتظهر النتائج كما هو موضح في الشكل 19):

الشكل 19

	A	B	C
1	t-Test: Paired Two Sample for Means		
2			
3		قبل استعمال العقار	بعد شهر من استعمال العقار
4	Mean	84.5	81.25
5	Variance	92.63636364	84.56818182
6	Observations	12	12
7	Pearson Correlation	0.989611623	
8	Hypothesized Mean Difference	0	
9	df	11	
10	t Stat	7.915991495	
11	P(T<=t) one-tail	3.61045E-06	
12	t Critical one-tail	1.795884819	
13	P(T<=t) two-tail	7.22089E-06	
14	t Critical two-tail	2.20098516	

قيمة معامل الارتباط بيرسون (سنتكلم عنه لاحقا)

معنوية الاختبار حالة الفرضية البديلة أحادية الذيل

معنوية الاختبار حالة الفرضية البديلة ثنائية الذيل

يتضح من قيمة معنوية الاختبار أحادية الذيل P(T<=t) one-tail وهي تقريبا 0.0000 والتي تصغر مستوى الدلالة 0.05 أننا نرفض الفرضية الابتدائية القائلة بتساوي المتوسطين، ونقبل الفرضية البديلة القائلة بأن متوسط وزن النساء بعد العقار أكبر من وزنهن قبل العقار. بمعنى أن للعقار اثرا معنويا ملحوظا على تخفيض وزن النساء في غضون شهر.

ناقشنا في الفصل الثاني المقارنات المتعلقة بمجتمعين على الأكثر، لكن ماذا لو كان لدينا ثلاث مجموعات بأكثر. للتوضيح، أرادت شركة تصنيع سيارات اختبار تأثير درجة الحرارة على درجة تماسك طلاء السيارات الخارجي مع تغير درجات الحرارة. فأخذنا عينتين عشوائيتين

من السيارات وطبقنا على كل عينة درجة حراراة مختلفة، ثم قارنا بين متوسطي درجة تماسك كل مجموعة، وبناء عليه، نقرر فيما إذا كان هنالك فرق معنوي بين المجموعتين أم لا (أي هل يوجد تأثير لدرجة الحرارة على تماسك الطلاء أم لا). نلاحظ هنا حساسية القرار بناء على درجة الحرارة التي أعطيت لكل مجموعة، فلو افترضنا مثلا أنه أجرينا الاختبار ونتج أنه لا يوجد تاثير لدرجة الحرارة على تماسك الطلاء، ثم أعدنا الاختبار لكن مع تطبيق درجة حرارة مختلفة لإحدى المجموعتين. هنا قد نحصل على نتائج مختلفة. إذا نحن هنا أمام حالة لا يكفيها التقسيم إلى مجموعتين، بل نحتاج إلى تقسيم مجتمع السيارات إلى ثلاثة مجموعات على الأقل بحيث نطبق على كل مجموعة درجة حرارة مختلفة (منخفضة، متوسطة، مرتفعة)، وثم دراسة الفروقات بين المتوسطات الثلاثة والتقرير فيما إذا كان هناك تأثير معنوي لدرجة الحرارة على تماسك الطلاء أم لا. لدراسة هذه القضية سنحتاج إلى استخدام تحليل التباين.

3. تحليل التباين Analysis Of Variance (ANOVA)

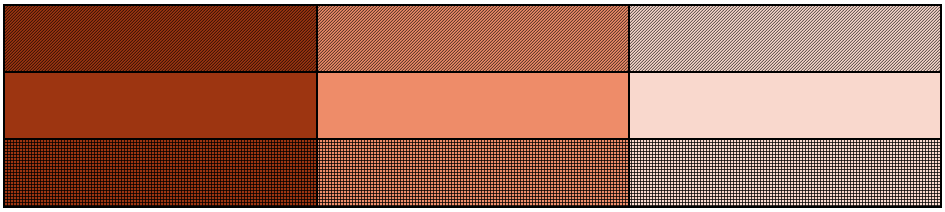
عرضنا سابقا مثال دراسة تأثير درجة الحرارة على تماسك طلاء السيارة، حينها اقترحنا تقسيم العينة إلى ثلاثة مجموعات متشابهة من السيارات، ومن ثم تطبيق درجات حرارة مختلفة على كل مجموعة، لدراسة تأثير درجة الحرارة على تماسك الطلاء نستخدم هنا تحليل التباين أحادي الاتجاه One-Way Analysis of Variance. كما يوضح الشكل 20 الآتي:

الشكل 20

مجموعة سيارات طبقت عليها درجة حرارة منخفضة	مجموعة سيارات طبقت عليها درجة حرارة متوسطة	مجموعة سيارات طبقت عليها درجة حرارة مرتفعة

لنفرض هنا أنن نريد اختبار تأثير ثلاثة مستويات متدرجة لدرجة الحرارة على ثلاثة أنواع مختلفة من الطلاء (لنقل من ثلاثة موردين مختلفين). هنا قمنا مرة ثانية بتقسيم كل مجموعة من السيارات التي ستخضع لنفس درجة الحرارة إلى ثلاثة مجموعات جديدة كل مجموعة لمورد طلاء مختلف، كما يوضح الشكل 21:

الشكل 21



هذه الحالة التي تتضمن التقسيم الثنائي نستخدم معه تحليل التباين ثنائي الاتجاه Two-Way Analysis of Variance.

1.3 تحليل التباين أحادي الاتجاه:

سنوضح هذا الاختبار من خلال التطبيق الآتي:

أرادت شركة سيارات تطوير التصميم الخارجي لإحدى سياراتها، فطلبت من أربع مصممين تقديم التصميم الجديد المقترح، ثم اجتمع مدراء الإنتاج في الشركة لتقييم التصاميم الجديدة، فكانت تقييماتهم كالتالي (التقييم من 100):

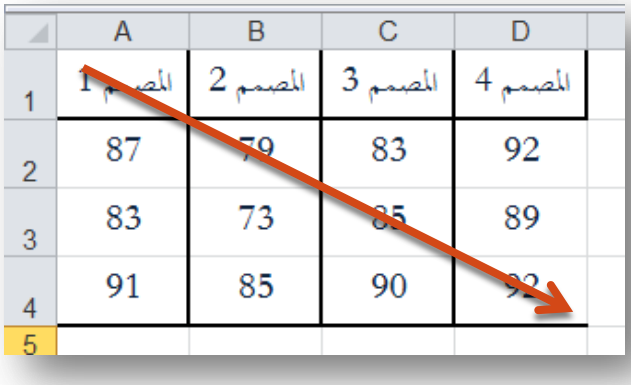
	المصمم			
	1	2	3	4
تقييمات مدراء الإنتاج	87 83 91	79 73 85	83 85 90	92 89 92
متوسط التقييم للمصمم	87	79	86	91

1- هل الاختلاف بين التقييمات ناتج عن اختلاف مهارات المصممين فعلا، أم الاختلاف ناتج عن اختلاف أذواق وآراء المدراء ليس إلا؟

2- بملاحظة أن هناك فرقا بين تقييمات المصممين، هل هذا الاختلاف جوهري؟

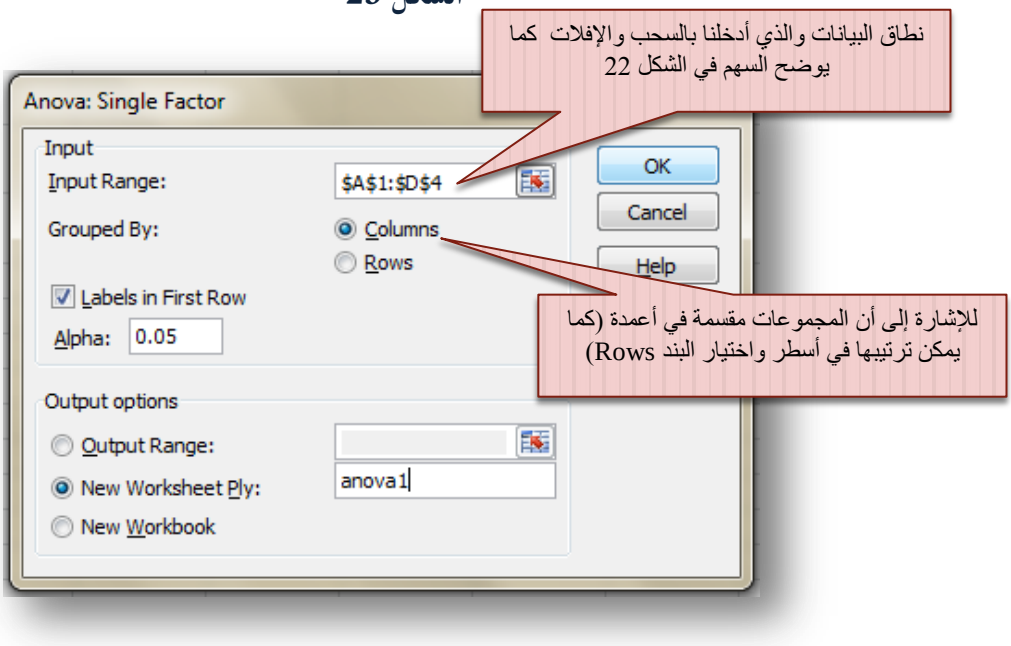
سندخل البيانات في برنامج إكسل كما يوضح الشكل 22 الآتي:

الشكل 22



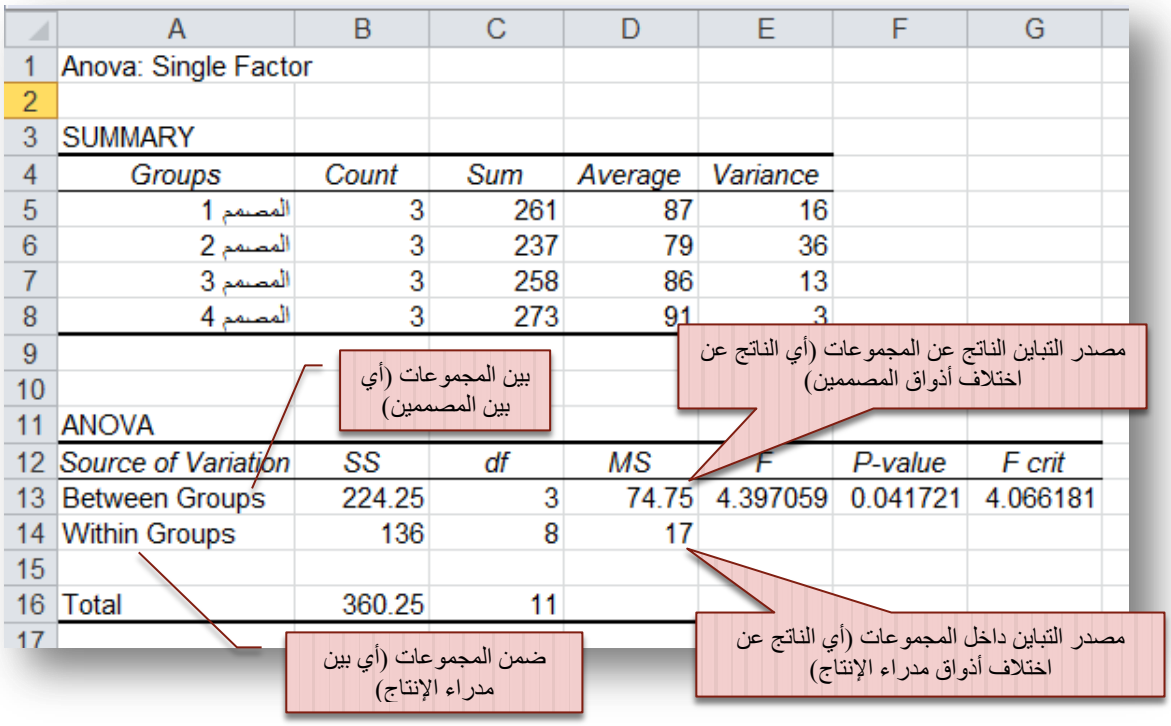
الإجابة على الطلبين الأول والثاني يتطلب منا إنشاء جدول تحليل التباين بحيث نعتبر أن المجموعات هي المصممين، بمعنى أن لكل مصمم مجموعته الخاصة من تقييمات مدراء الإنتاج لتصميمه، لذا نختار الوظيفة Anova Single Factor من الأداة Data Analysis فتظهر لدينا النافذة الآتية (الشكل 23):

الشكل 23



فتظهر لدينا النتائج الآتية (الشكل 24):

الشكل 24



	A	B	C	D	E	F	G
1	Anova: Single Factor						
2							
3	SUMMARY						
4	*Groups*	*Count*	*Sum*	*Average*	*Variance*		
5	المصمم 1	3	261	87	16		
6	المصمم 2	3	237	79	36		
7	المصمم 3	3	258	86	13		
8	المصمم 4	3	273	91	3		
9							
10							
11	ANOVA						
12	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
13	Between Groups	224.25	3	74.75	4.397059	0.041721	4.066181
14	Within Groups	136	8	17			
15							
16	Total	360.25	11				
17							

تفسير مخرجات جدول تحليل التباين:

- نلاحظ أن أفضل مصمم هو المصمم 4 الذي حصل على متوسط تقييم 91، وأسوؤهم هو المصمم 2 الذي حصل على تقييم 79.
- نلاحظ أن التباين الناتج عن اختلاف أذواق (أو مهارات) المصممين 74.75، أكبر من التباين الناتج عن اختلاف آراء مدراء الإنتاج 17. بمعنى أن هناك اختلافا كبير فعلا بين مهارات المصممين.
- من الواضح أن هناك اختلافا بين المصممين (MS Between Groups) وهو أكبر من الاختلاف الناتج عن مدراء الإنتاج (MS Within Groups)، لكن بالنظر إلى قيمة معنوية الاختبار (P-value) التي تساوي 0.0417 نجد أنها أصغر من مستوى الدلالة 0.05 الذي يؤدي إلى رفض الفرضية الابتدائية القائلة بتساوي متوسطات تقييم المصممين، وقبول البديلة القائلة بان هناك اختلافا معنويا بين مهارات المصممين.

تنويه هام: تستطيع استخدام تحليل التباين أحادي الاتجاه حتى في حالة عدم تساوي المجموعات.

2.3 تحليل التباين ثنائي الاتجاه:

سنوضح هذا الاختبار من خلال التطبيق الآتي:

أراد مدير التسويق في شركة ما اختبار أثر كل من الإعلانات والتخفيضات على المبيعات الأسبوعية لأحد منتجات الشركة، فأخذت مبيعات أسبوعين من كل حالة، فكانت النتائج:

التسعير	مع حملة إعلانية	بدون حملة إعلانية

مع تخفيض	9.8	6
مع تخفيض	10.6	5.3
بدون تخفيض	6.2	4.3
بدون تخفيض	7.1	3.9

1. أي تقنيات التسويق كان لها تأثير أكبر على المبيعات: التخفيضات أم الإعلانات؟
2. هل هناك تأثير معنوي للتخفيضات على المبيعات؟
3. هل هناك تأثير معنوي للحملات الإعلانية على المبيعات؟

بمعنى أنه لدينا الفرضيتين الآتين:

- **الفرضية الأولى**: تقول الفرضية الابتدائية أنه لا يوجد تاثير معنوي للتخفيضات على المبيعات (متوسط مبيعات الأسابيع التي لم تتضمن تخفيضات تساوي متوسط مبيعات الأسابيع مع تخفيضات) مقابل الفرضية البديلة القائلة بأنه يوجد تاثير معنوي للتخفيضات على المبيعات.
- **الفرضية الثانية**: تقول الفرضية الابتدائية أنه لا يوجد تاثير معنوي للحملات الإعلانية على المبيعات (متوسط مبيعات الأسابيع التي لم تتضمن حملات الإعلانية تساوي متوسط مبيعات الأسابيع مع حملات الإعلانية) مقابل الفرضية البديلة القائلة بأنه يوجد تاثير معنوي للحملات الإعلانية على المبيعات.

للإجابة على هذه الأسئلة سنستخدم تحليل التباين ثنائي الاتجاه، من أجل ذلك أدخل البيانات في برنامج إكسل كما هو موضح في الشكل 25:

الشكل 25

	A	B	C
1	التسعير	مع حملة إعلانية	بدون حملة إعلانية
2	مع تخفيض	9.8	6
3	مع تخفيض	10.6	5.3
4	بدون تخفيض	6.2	4.3
5	بدون تخفيض	7.1	3.9
6			

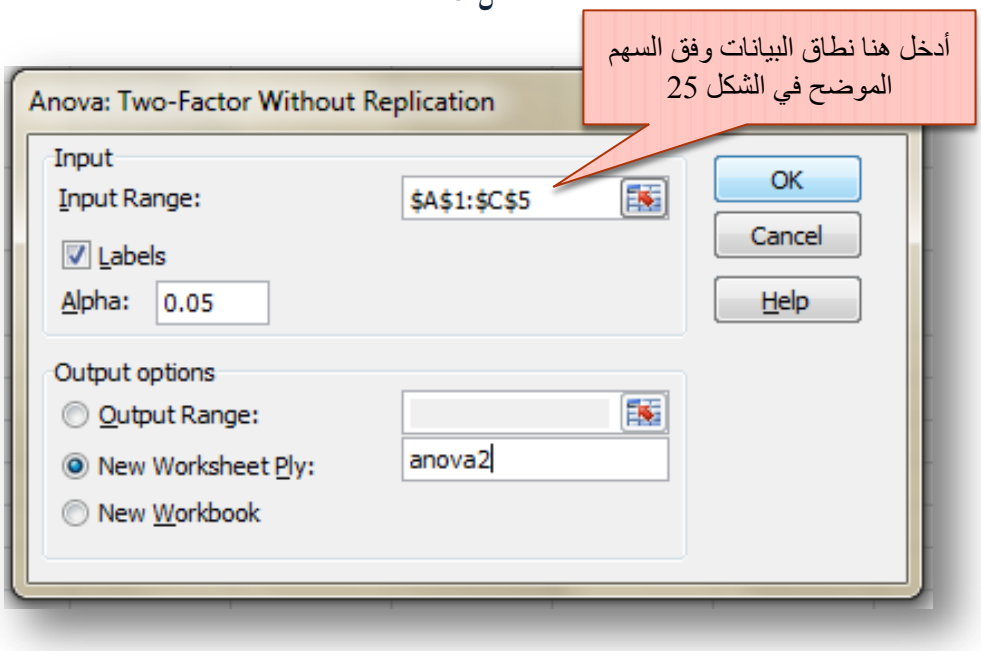
- لاحظ أن عمود التسعير وضع على اليسار، وذلك بسبب أن اتجاه الصفحة هو من اليسار إلى اليمين، وإلا فيجب أن يكون العمود من اليمين (كما يوضح الشكل 26).
- لاحظ تكرار عبارة (مع تخفيض) و(بدون تخفيض) في كل سطر.

الشكل 26

	A	B	C
1	التسعير	مع حملة إعلانية	بدون حملة إعلانية
2	مع تخفيض	9.8	6
3	مع تخفيض	10.6	5.3
4	بدون تخفيض	6.2	4.3
5	بدون تخفيض	7.1	3.9
6			

لذا نختار الوظيفة Anova: Two Factor Without Replication من الأداة Data Analysis فتظهر لدينا النافذة الآتية (الشكل 27):

الشكل 27



فتظهر لدينا النتائج الموضحة في الشكل 28:

الشكل 28

	A	B	C	D	E	F	G
1	Anova: Two-Factor Without Replication						
2							
3	*SUMMARY*	*Count*	*Sum*	*Average*	*Variance*		
4	مع تخفيض	2	15.8	7.9	7.22		
5	مع تخفيض	2	15.9	7.95	14.045		
6	بدون تخفيض	2	10.5	5.25	1.805		
7	بدون تخفيض	2	11	5.5	5.12		
8							
9	مع حملة إعلانية	4	33.7	8.425	4.4425		
10	بدون حملة إعلانية	4	19.5	4.875	0.909167		
11							
12							
13	ANOVA						
14	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
15	Rows	13.07	3	4.356667	4.378559	0.128234	9.276628
16	Columns	25.205	1	25.205	25.33166	0.015117	10.12796
17	Error	2.985	3	0.995			
18							
19	Total	41.26	7				
20							
21							
22							

متوسط التباين الناتج عن التخفيضات

متوسط التباين الناتج عن الإعلانات

تفسير المخرجات:

- يقصد بالأسطر rows دراسة تأثير التخفيض، بينما يقصد بالأعمدة Columns دراسة تأير الحملات الإعلانية. الشكل 29 يوضح ذلك.
- واضح أن متوسط التباين الناتج عن الإعلانات أكبر من متوسط التباين الناتج عن التخفيضات، بمعنى أن الإعلانات قد أحدثت أثرا أكبر (اختلافات أكبر بين المجموعات) على المبيعات من التخفيضات.
- يتضح من قيمة المعنوية المقابلة للتخفيضات والتي تساوي 0.128 أنها أكبر من مستوى الدلالة 0.05 وبالتالي فإننا نقبل الفرضية الابتدائية القائلة بعدم وجود تأثير للتخفيضات، بمعنى أن متوسط مبيعات الأسابيع التي لم تتضمن تخفيضات تساوي متوسط مبيعات الأسابيع مع تخفيضات عند مستوى دلالة 0.05.
- لكن بالمقابل فإن قيمة المعنوية المقابلة للحملات الإعلانية والتي تساوي 0.015 أنها أصغر من مستوى الدلالة 0.05 وبالتالي فإننا نرفض الفرضية الابتدائية القائلة بعدم وجود تأثير للحملات الإعلانية، بمعنى أن متوسط مبيعات الأسابيع التي لم تتضمن حملات الإعلانية لا تساوي متوسط مبيعات الأسابيع مع حملات الإعلانية عند مستوى دلالة 0.05.
- الخلاصة: يوجد تأثير معنوي للحملات الإعلانية على المبيعات ولا يوجد تأثير معنوي للتخفيضات. يوضح الشكل 30 مخطط أعمدة ثلاثي الأبعاد للمبيعات لكل أسبوع، أما الشكل 31 فيوضح متوسط مبيعات كل تقسيم.

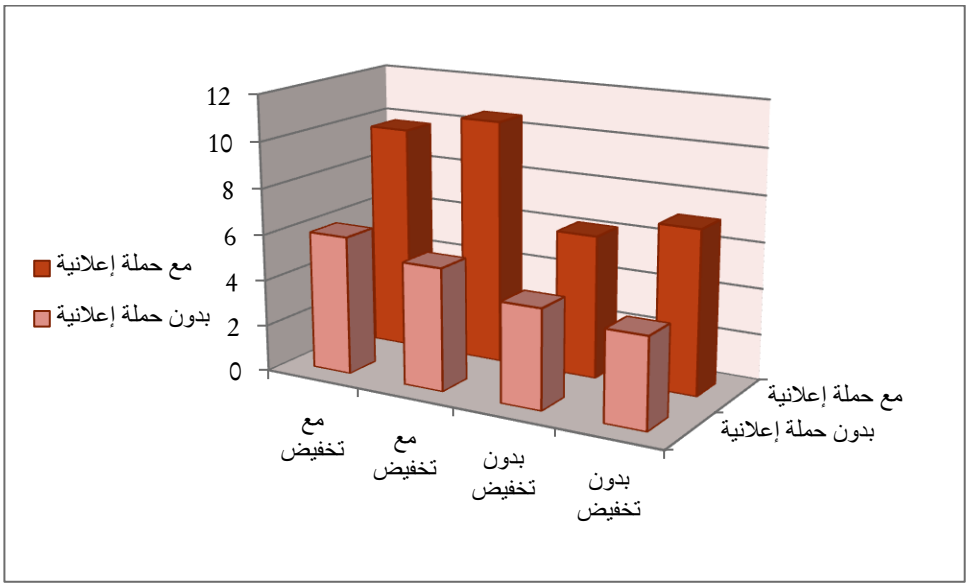
الشكل 29

	A	B	C
1	التسعير	مع حملة إعلانية	بدون حملة إعلانية
2	مع تخفيض	9.8	6
3	مع تخفيض	10.6	5.3
4	بدون تخفيض	6.2	4.3
5	بدون تخفيض	7.1	3.9
6			

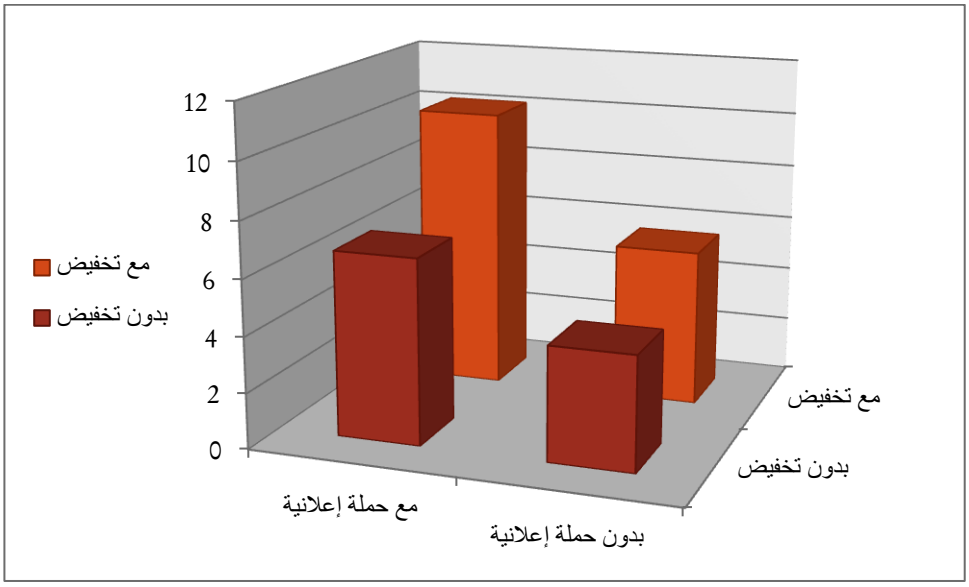
يتضمن كل سطر حالتين:
مع وبدون حملة إعلانية

يتضمن كل عمود حالتين:
مع وبدون تخفيضات

الشكل 30



الشكل 31



واضح من الشكل 31 أن متوسط حجم المبيعات عند تطبيق كل من حملة التخفيضات مع الحملة الإعلانية له أكبر قيمة (أطول عمود كما هو واضح)، فهل هذا العمود هو أكبر من بقية الأعمدة لدرجة كبيرة؟ بعبارة أخرى، هل كان للجمع بين نشاطين تسويقيين وهما التخفيضات والحملة الإعلانية أثرا معنويا مختلفا عن استخدام إحدى النشاطين التسويقيين فقط؟

الحقيقة أن تحليل التباين ثنائي الاتجاه بدون إعادة (ANOVA Two-Factor without Replication) لا يجيبنا على السؤال، أي أنه لا يتضمن دراسة ما نسميه "التفاعل" بين عاملين. وفي مثالنا دراسة التفاعل بين كل من التخفيضيات والحملات الإعلانية. سنستعرض الآن تحليل التباين ثنائي الاتجاه تام العشوائية (ويسمى مع إعادة أيضا) والذي يتضمن دراسة معنوية: عامل **أ**، عامل **ب**، عامل **أ*ب** (أي دراسة التفاعل بين العاملين **أ** و **ب**).

3.3 تحليل التباين ثنائي الاتجاه تام العشوائية:

قامت شركة بحوث تسويق باختبار رأي المستهلكين لشامبو جديد على أربع فئات عمرية مختلفة في أربع مناطق جغرافية مختلفة أيضا. كما يوضح الجدول الآتي. في كل تصنيف، سالت الشركة خمسة أشخاص بحيث سؤل كل واحد منهم عشرة أسئلة حول رضاهم عن الشامبو الجديد من عدة جوانب مختلفة (قيم كل جانب بدرجة من 5)، ثم سجلت متوسط درجات رضا كل فرد (أي يدل المتوسط على إجمالي الرضا العام):

	المنطقة الشمالية	المنطقة الجنوبية	المنطقة الوسطى	المنطقة الغربية
الشباب (تحت 18)	3.9	3.9	3.6	3.9
	4.0	4.2	3.9	4.4
	3.7	4.4	3.9	4.0
	4.1	4.1	3.7	4.1
	4.3	4.0	3.3	3.9
الجامعيين (18-25)	4.0	3.8	3.6	3.8
	4.0	3.7	4.1	3.8
	3.7	3.7	3.8	3.6
	3.8	3.6	3.9	3.6
	3.8	3.7	4.0	4.1
البالغين (26-64)	3.2	3.5	3.5	3.8
	3.8	3.3	3.8	3.6
	3.7	3.4	3.8	3.4
	3.4	3.5	4.0	3.7
	3.4	3.4	3.7	3.1
الشيوخ (أكبر من 65)	3.4	3.6	3.3	3.4
	2.9	3.4	3.3	3.2
	3.6	3.6	3.1	3.5
	3.7	3.6	3.1	3.3
	3.5	3.4	3.1	3.4

الأسئلة المطروحة هي:

- هل هناك فرق بين رضا الفئات العمرية؟ (أي هل هناك تأثير معنوي لعامل العمر على الرضا؟)
- هل هناك فرق بين رضا المناطق الجغرافية؟ (أي هل هناك تأثير معنوي لعامل المنطقة على الرضا؟)
- هل يرتبط تأثير المنطقة الجغرافية بفئة عمرية محددة؟ أو هل يرتبط تأثير الفئة العمرية على منطقة جغرافية محددة؟ (بمعنى هل هناك تفاعل بين عاملي العمر والمنطقة؟)

والفرضيات المقابلة لها هي:

- لا يوجد تأثير معنوي للعمر على الرضا – مقابل: يوجد تأثير معنوي للعمر على الرضا
- لا يوجد تأثير للمنطقة الجغرافية على الرضا – مقابل: يوجد تأثير للمنطقة الجغرافية على الرضا
- لا يوجد تفاعل بين العمر والمنطقة الجغرافية – مقابل: يوجد تفاعل بين العمر والمنطقة الجغرافية

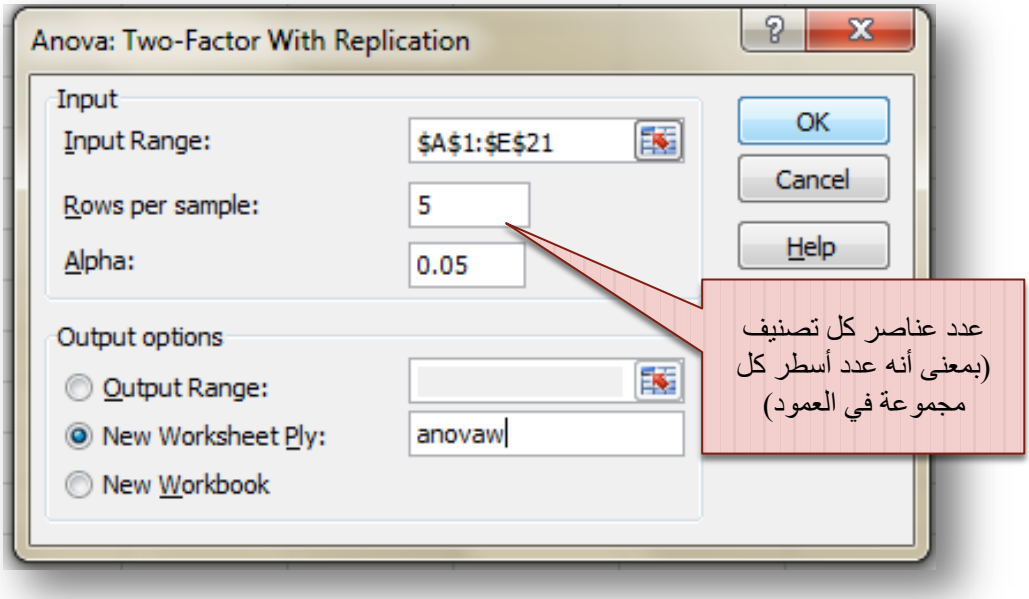
لندخل أولا البيانات في إكسل كما يوضح الشكل 32 الآتي:

الشكل 32

	A	B	C	D	E
1		المنطقة الشمالية	المنطقة الجنوبية	المنطقة الوسطي	المنطقة الغربية
2	الشباب (تحت 18)	3.9	3.9	3.6	3.9
3	الشباب (تحت 18)	4.0	4.2	3.9	4.4
4	الشباب (تحت 18)	3.7	4.4	3.9	4.0
5	الشباب (تحت 18)	4.1	4.1	3.7	4.1
6	الشباب (تحت 18)	4.3	4.0	3.3	3.9
7	الجامعيين (18-25)	4.0	3.8	3.6	3.8
8	الجامعيين (18-25)	4.0	3.7	4.1	3.8
9	الجامعيين (18-25)	3.7	3.7	3.8	3.6
10	الجامعيين (18-25)	3.8	3.6	3.9	3.6
11	الجامعيين (18-25)	3.8	3.7	4.0	4.1
12	البالغين (26-64)	3.2	3.5	3.5	3.8
13	البالغين (26-64)	3.8	3.3	3.8	3.6
14	البالغين (26-64)	3.7	3.4	3.8	3.4
15	البالغين (26-64)	3.4	3.5	4.0	3.7
16	البالغين (26-64)	3.4	3.4	3.7	3.1
17	الشيوخ (أكبر من 65)	3.4	3.6	3.3	3.4
18	الشيوخ (أكبر من 65)	2.9	3.4	3.3	3.2
19	الشيوخ (أكبر من 65)	3.6	3.6	3.1	3.5
20	الشيوخ (أكبر من 65)	3.7	3.6	3.1	3.3
21	الشيوخ (أكبر من 65)	3.5	3.4	3.1	3.4
22					

اختر الوظيفة Anova: Two Factor With Replication من الأداة Data Analysis فتظهر لدينا النافذة الآتية (الشكل 33):

الشكل (33)



لاحظ أنك تستطيع إعادة ترتيب البيانات في صفحة إكسل كما يوضحه الشكل 34، والإدخال لن يختلف.

الشكل 34

	A	B	C	D	E	F
1		الشباب (تحت 18)	الجامعيين (18-25)	البالغين (26-64)	الشيوخ (أكبر من 65)	
2	المنطقة الشمالية	3.9	4.0	3.2	3.4	
3	المنطقة الشمالية	4.0	4.0	3.8	2.9	
4	المنطقة الشمالية	3.7	3.7	3.7	3.6	
5	المنطقة الشمالية	4.1	3.8	3.4	3.7	
6	المنطقة الشمالية	4.3	3.8	3.4	3.5	
7	المنطقة الجنوبية	3.9	3.8	3.5	3.6	
8	المنطقة الجنوبية	4.2	3.7	3.3	3.4	
9	المنطقة الجنوبية	4.4	3.7	3.4	3.6	
10	المنطقة الجنوبية	4.1	3.6	3.5	3.6	
11	المنطقة الجنوبية	4.0	3.7	3.4	3.4	
12	المنطقة الغربية	3.9	3.8	3.8	3.4	
13	المنطقة الغربية	4.4	3.8	3.6	3.2	
14	المنطقة الغربية	4.0	3.6	3.4	3.5	
15	المنطقة الغربية	4.1	3.6	3.7	3.3	
16	المنطقة الغربية	3.9	4.1	3.1	3.4	
17	المنطقة الوسطي	3.6	3.6	3.5	3.3	
18	المنطقة الوسطي	3.9	4.1	3.8	3.3	
19	المنطقة الوسطي	3.9	3.8	3.8	3.1	
20	المنطقة الوسطي	3.7	3.9	4.0	3.1	
21	المنطقة الوسطي	3.3	4.0	3.7	3.1	
22						

لكن سندخل البيانات كما يوضحه الشكل 33 (بناء على البيانات المرتبة كما في الشكل 32، أي أن الأعمدة تمثل عامل المنطقة، والأسطر تمثل عامل الفئة العمرية) تظهر لدينا المخرجات الآتية الشكل 35 (أ وب):

الشكل 35 – أ

	A	B	C	D	E	F
1	Anova: Two-Factor With Replication					
2						
3	SUMMARY	المنطقة الشمالية	المنطقة الجنوبية	المنطقة الوسطى	المنطقة الغربية	Total
4	الشباب (تحت 18)					
5	Count	5	5	5	5	20
6	Sum	20	20.6	18.4	20.3	79.3
7	Average	4	4.12	3.68	4.06	3.965
8	Variance	0.05	0.037	0.062	0.043	0.070816
9						
10	الجامعيين (18-25)					
11	Count	5	5	5	5	20
12	Sum	19.3	18.5	19.4	18.9	76.1
13	Average	3.86	3.7	3.88	3.78	3.805
14	Variance	0.018	0.005	0.037	0.042	0.026816
15						
16	البالغين (26-64)					
17	Count	5	5	5	5	20
18	Sum	17.5	17.1	18.8	17.6	71
19	Average	3.5	3.42	3.76	3.52	3.55
20	Variance	0.06	0.007	0.033	0.077	0.054211
21						
22	الشيوخ (أكبر من 65)					
23	Count	5	5	5	5	20
24	Sum	17.1	17.6	15.9	16.8	67.4
25	Average	3.42	3.52	3.18	3.36	3.37
26	Variance	0.097	0.012	0.012	0.013	0.044316
27						
28	Total					
29	Count	20	20	20	20	
30	Sum	73.9	73.8	72.5	73.6	
31	Average	3.695	3.69	3.625	3.68	
32	Variance	0.1089211	0.0883158	0.1051316	0.1111579	
33						

متوسط رضا الشباب الإجمالي

متوسط رضا الجامعيين في المنطقة الغربية

متوسط رضا البالغين في كل المناطق

متوسط رضا البالغين في المنطقة الوسطى

متوسط رضا المنطقة الجنوبية

توضح هذه المخرجات الإحصاء الوصفي للحالة المدروسة، أما جدول تحليل التباين فيوضحه الشكل 35 – ب الآتي:

الشكل 35– ب

	A	B	C	D	E	F	G
34							
35	ANOVA						
36	*ırce of Variat*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
37	Sample	4.1925	3	1.3975	36.958678	5.56E-14	2.748191
38	Columns	0.0625	3	0.0208333	0.5509642	0.649331	2.748191
39	Interaction	1.2445	9	0.1382778	3.656933	0.000977	2.029792
40	Within	2.42	64	0.0378125			
41							
42	Total	7.9195	79				

معنوية تأثير عامل العمر

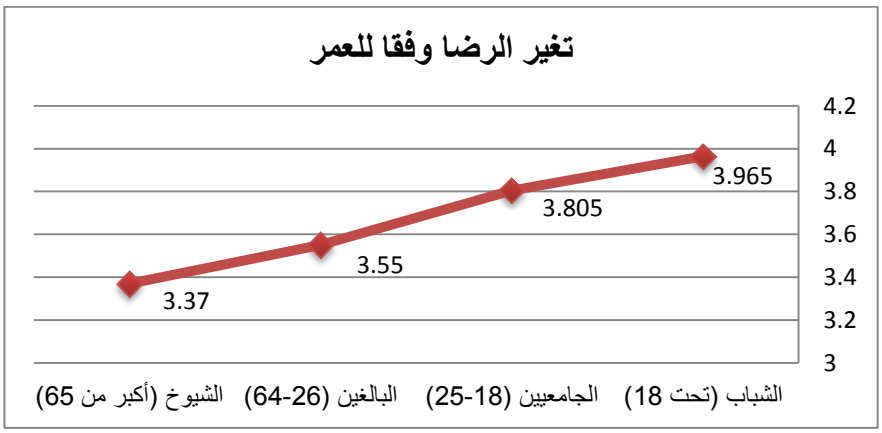
معنوية تأثير عامل المنطقة الجغرافية

معنوية تأثير تفاعل الفئة العمرية مع المنطقة

تفسير المخرجات:

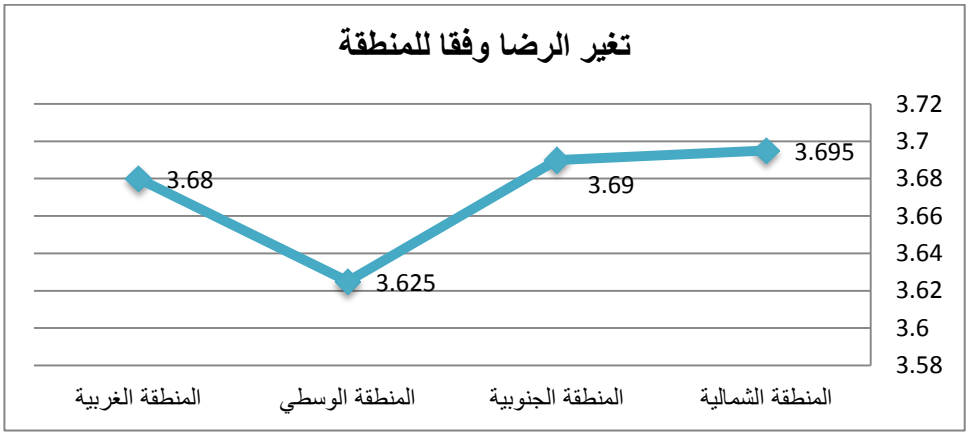
- يؤثر العمر معنويا على درجات الرضا (إذ تبلغ المعنوية الصفر تقريبا 0.000 وهي أصغر من مستوى الدلالة 0.05)، بحيث يرتفع عند الشباب، ويتدرج نزولا حتى فئة الشيوخ الأقل رضا كما يوضح الشكل 36:

الشكل 36



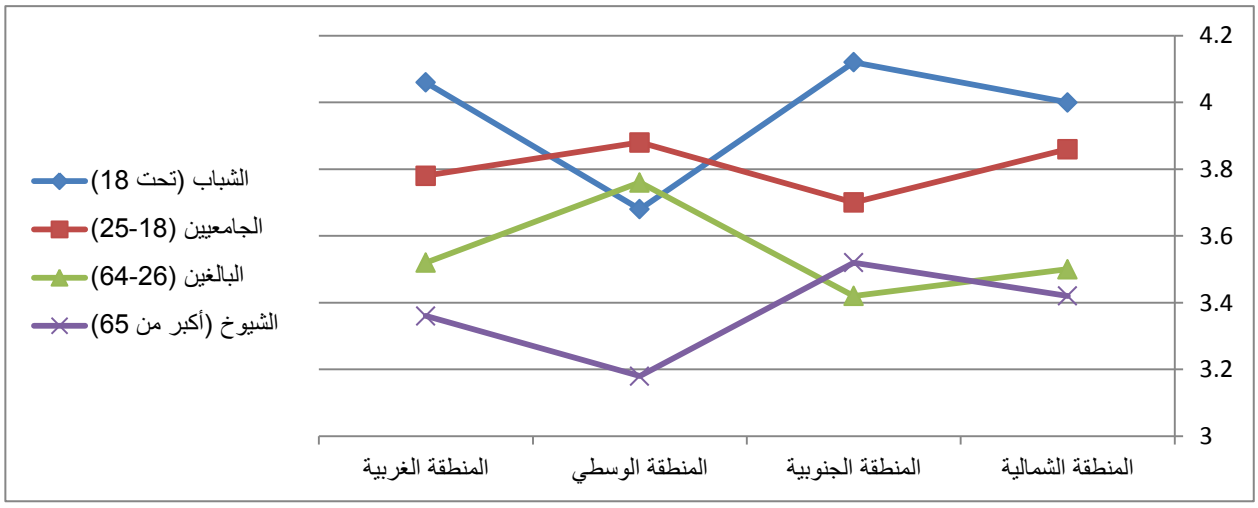
- لا تؤثر المنطقة الجغرافية على الرضا (إذ تبلغ المعنوية 0.649 أكبر من 0.05)، إذ يتساوى تقريبا في كل المناط باستثناء المنطقة الوسطى التي شهدت انخفاضا طفيفا في الرضا. انظر الشكل 37.

الشكل 37

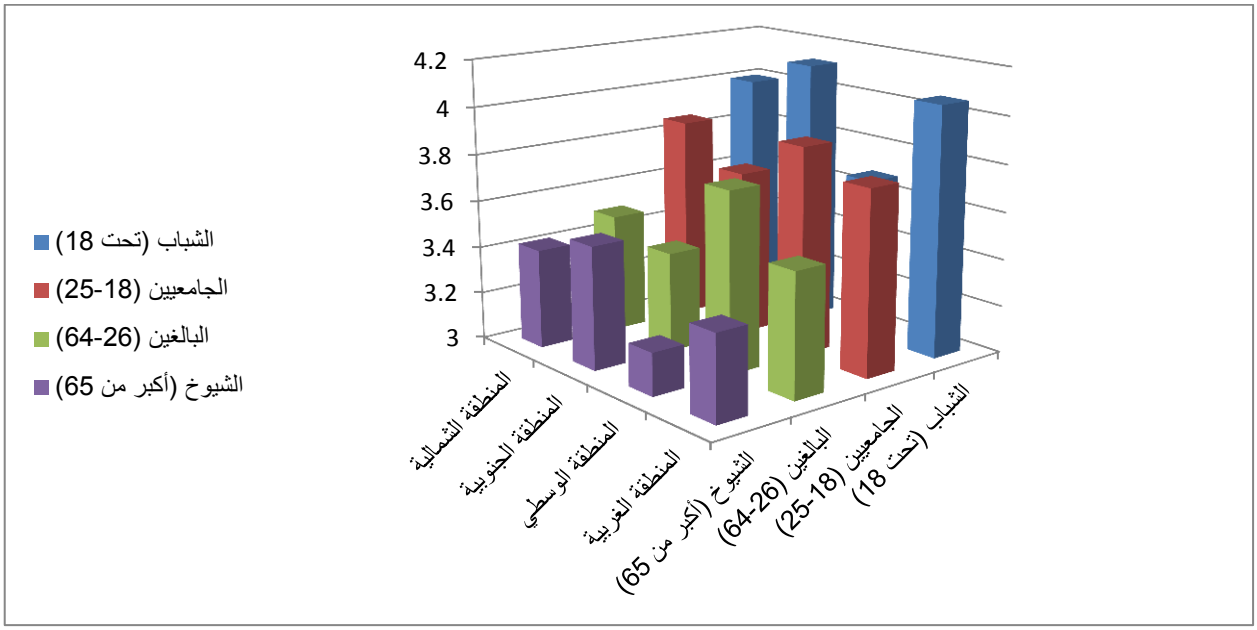


يوجد تفاعل معنوي بين المنطقة والفئة العمرية (حيث تبلغ معنويتها 0.001 وهي أصغر من مستوى الدلالة 0.05)، نلاحظ أن ارتفاع الرضا عند فئتي الجامعيين والبالغين يرتبط بكونهم يقطنون في المنطقة الوسطى. بمعنى أن هناك عوامل ما تؤثر على هتين الفئتين موجودة فقط في المنطقة الوسطى أدت إلى ارتفاع الرضا عند كل من الجامعيين والبالغين. يوضح الشكل 38 الخطوط الممثلة لمتوسطات التصنيفات، أو الشكل 39 الذي يمثل المتوسطات بالأعمدة، لاحظ عند تعدد التصنيفات يصبح مخطط الأعمدة صعب القراءة على خلال مخطط الخطوط الذي تسهل قراءته دائما.

الشكل 38: تفاعل المنطقة مع الفئة العمرية



الشكل 39



4. دراسة الارتباط وتحليل الانحدار

Regression & Correlation Analysis

استعرضنا حتى الآن ما يسمى مقارنة المتوسطات بين مجموعتين أو أكثر. سنستعرض الآن كيفية دراسة العلاقة بين متحول واحد أو عدة متحولات (نسميهم المتحولات المستقلة Independent Variables أو المتحولات المنبئة Predictors) من جهة، بين متحول تابع Dependent وحيد من جهة أخرى.

الطرق التي سنتعتمدها في دراسة هذه العلاقة هي معامل الارتباط بيرسون Pearson Correlation Coefficient وتحليل الانحدار Regression Analysis. فإذا كان لدينا متحول مستقل وحيد نتعامل مع ما يسمى "الانحدار البسيط" و"معامل الارتباط البسيط"، وإذا كان لدينا أكثر من متحول مستقل نستخدم ما يسمى "الانحدار المتعدد" و"معامل الارتباط المتعدد".

1.4 معامل الارتباط البسيط بيرسون:

يقيس معامل الارتباط بيرسون درجة العلاقة بين ظاهرتين (متحولين كميين)، على سبيل المثال: هل هناك علاقة بين الطول والوزن؟ للإجابة على هذا السؤال يمكن أن نأخذ عينة عشوائية من الأفراد نسجل لكل فرد ثنائية من المعلومات تضم كل ثنائية: (الطول، الوزن)، ومن ثم نحسب معامل الا رتباط بيرسون لهم، والذي تنحصر قيم هذا المعامل في المجال المغلق: [-1, +1] ونقرر شكل العلاقة بين الوزن والطول وفقا لقيمة المعامل:

قيمة المعامل	شكل العلاقة
1 تماما	علاقة طردية خطية تامة
بين 0.9 و 1	علاقة طردية شبه خطية
بين 0.6 و 0.9	علاقة طردية قوية
بين 0.4 و 0.6	علاقة طردية متوسطة
من 0 حتى 0.4	علاقة طردية ضعيفة
0 تماما	لا توجد علاقة أبدا
بين 0 و 0.4-	علاقة عكسية ضعيفة
بين 0.4- و 0.6-	علاقة عكسية متوسطة
بين 0.6- و 0.9-	علاقة عكسية قوية
بين 0.9- و 1-	علاقة عكسية طردية شبه خطية
1- تماما	علاقة عكسية طردية خطية تامة

فإذا كانت قيمة المعامل الناتج مثلا 0.65 فإننا نقول أن العلاقة بين الطول والوزن طردية متوسطة، وكلمة "طردية" تعني أنه مع **زيادة** الطول **سيزداد** الوزن باحتمال 65% (أو يمكن القول أن 65% من زيادة الوزن سببها زيادة الطول). أما إذا كان مثلا قيمة معامل الارتباط 0.85- فإننا نقول أن العلاقة بين الطول والوزن عكسية قوية، وكلمة "عكسية" تعني أنه مع **زيادة** الطول **سينقص** الوزن باحتمال 85% (أو يمكن القول أن 85% من زيادة الوزن سببها نقصان الطول). بالطبع، هذه النتائج يمكن تفسيرها بطريقة أخرى، إذ نستطيع أن القول أن 85% من زيادة الطول سببها نقصان الوزن. بمعنى أن معامل الارتباط لا يفترض وجود متحول سبب ومتحول نتيجة، أي أن معامل الارتباط يدرس العلاقة المتبادلة بين متحولين متكافئين. بينما يتعدى تحليل الانحدار إلى دراسة العلاقة بين سبب ونتيجة، بمعنى أنه يدرس شكل ودرجة تأثير متحول سبب (مستقل) في متحول نتيجة (تابع).

لنأخذ المثال الآتي:

أردنا دراسة العلاقة بين تغير سعر سهمي شركة IBM وشركة EDS خلال شهري أكتوبر ونوفمبر عام 2004، فكانت البيانات موضحة في الجدول الآتي:

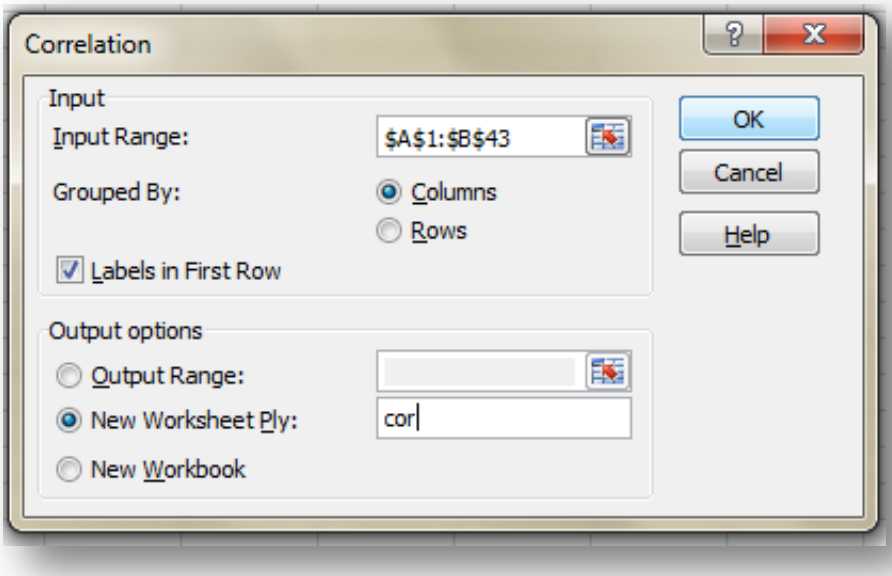
IBM	EDS	IBM	EDS
84.22	19.31	86.72	20.00
84.57	19.63	87.16	20.36
84.39	19.19	87.32	20.38
84.97	19.35	88.04	20.49
85.86	19.47	87.42	20.43
86.44	19.51	86.71	20.02
86.76	20.10	86.63	20.24
86.49	19.81	86.00	20.14
86.72	19.79	84.98	19.47
86.37	19.83	84.78	19.30
86.12	20.10	84.85	19.54
85.74	19.90	85.92	19.43
85.70	19.82	89.37	19.26
85.72	20.16	88.82	19.17
84.31	19.89	88.10	19.63
83.88	19.70	87.39	19.75
84.43	19.22	88.43	20.03
84.16	19.16	89.00	20.99
84.48	19.30	90.00	21.26
84.98	19.10	89.50	21.41
85.74	19.39	89.75	21.27

لدراسة العلاقة بين تغير سعر السهمين رتب البيانات في صفحة إكسل في عمودين وحيدين كما يوضح الشكل 40 الآتي:

الشكل 40

	A	B
1	IBM	EDS
2	84.22	19.31
3	84.57	19.63
4	84.39	19.19
5	84.97	19.35
6	85.86	19.47
7	86.44	19.51
8	86.76	20.10
9	86.49	19.81
10	86.72	19.79
11	86.37	19.83
12	86.12	20.10
13	85.74	19.90
14	85.70	19.82
15	85.72	20.16
16	84.31	19.89
17	83.88	19.70
18	84.43	19.22
19	84.16	19.16
20	84.48	19.30
21	84.98	19.10
22	85.74	19.39
23	86.72	20.00
24	87.16	20.36
25	87.32	20.38
26	88.04	20.49
27	87.42	20.43
28	86.71	20.02
29	86.63	20.24
30	86.00	20.14
31	84.98	19.47
32	84.78	19.30
33	84.85	19.54
34	85.92	19.43
35	89.37	19.26
36	88.82	19.17
37	88.10	19.63
38	87.39	19.75
39	88.43	20.03
40	89.00	20.99
41	90.00	21.26
42	89.50	21.41
43	89.75	21.27
44		

اختر الوظيفة Correlation من الأداة Data Analysis فتظهر لدينا النافذة الآتية (الشكل 41):



فتظهر لدينا المخرجات الآتية (الشكل 42):

الشكل 42

	A	B	C
1		*IBM*	*EDS*
2	IBM	1	
3	EDS	0.677054	1
4			
5			

يتضح من خلال معامل الارتباطبيرسون 0.68 أن هناك علاقة ارتباط طردية متوسطة، بمعنى أنه عند ارتفاع إحدى السهمين، فالسهم الآخر يتوقع ارتفاعه أيضا باحتمال 0.68

يعتبر معامل الارتباط بيرسون أبسط طرق اختبار شكل العلاقة بين متحولين، أما تحليل الانحدار فهو يفيد في اختبار شكل العلاقة إضافة إلى استخدامه في التنبؤ أيضا.

2.4 الانحدار البسيط:

كثيرا ما نلاحظ الظواهر الآتية:

- ترتبط المبيعات بنفقات الإعلانات
- يرتبط قسط التأجير الشهري لشقة ما بمساحتها
- ترتبط نسبة الوحدات المعيبة في خط إنتاج بسرعة الإنتاج

مثل هذه العلاقات يمكن نمذجتها بمعادلات، تتضمن كل معادلة متحولين (مستقل وتابع). على سبيل المثال لنفرض أننا توصلنا إلى أن المعادلة المناسبة لنمذجة المبيعات في شركة ما بالشكل: $Sales = 268 + 7.37Ad$

تجيبنا هذه المعادلة على السؤال العام: لو أنفقنا مبلغ Ad على الإعلانات، فإننا نتوقع أن نحصل على مبيعات قيمتها $Sales.

مثال: إذا أنفقت الشركة مبلغ 300 ألف$ على الإعلانات العام المقبل، فمن المتوقع أن الشركة ستحصل على عائد مبيعات حوالي: $Sales|_{Ad=300} = 268 + 7.37(300) = 2,449$ أي ستحصل الشركة على حوالي 2,449 ألف$ عند تطبيقها لحملة إعلانية بمبلغ 300 ألف$.

لكن كيف حصلنا على هذا النموذج؟ وإذا حصلنا عليه هل نتائجه دقيقة يعتمد عليها؟

تأخذ معادلة الانحدار الخطي البسيط الشكل العام: $Y=a+bX$

حيث نسمي Y بالمتحول التابع، ونسمي X بالمتحول المستقل، ونسمي a (intercept) بمعامل التقاطع، أما b فنسميه بالميل (slope).

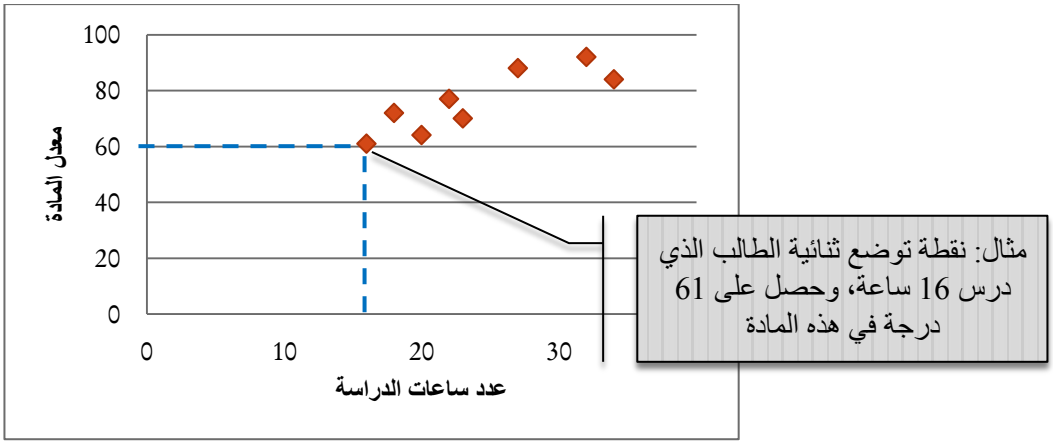
لنأخذ التطبيق الآتي:

لندرس تأثير متحول (عدد ساعات الدراسة في مادة ما) على نتيجة هذا الامتحان، كما يوضح الجدول الآتي:

رقم الطالب	1	2	3	4	5	6	7	8
عدد ساعات الدراسة (X)	20	16	34	23	27	32	18	22
معدل المادة (Y)	64	61	84	70	88	92	72	77

يوضح الشكل 43 لوحة الانتشار لزوج البيانات:

الشكل 43



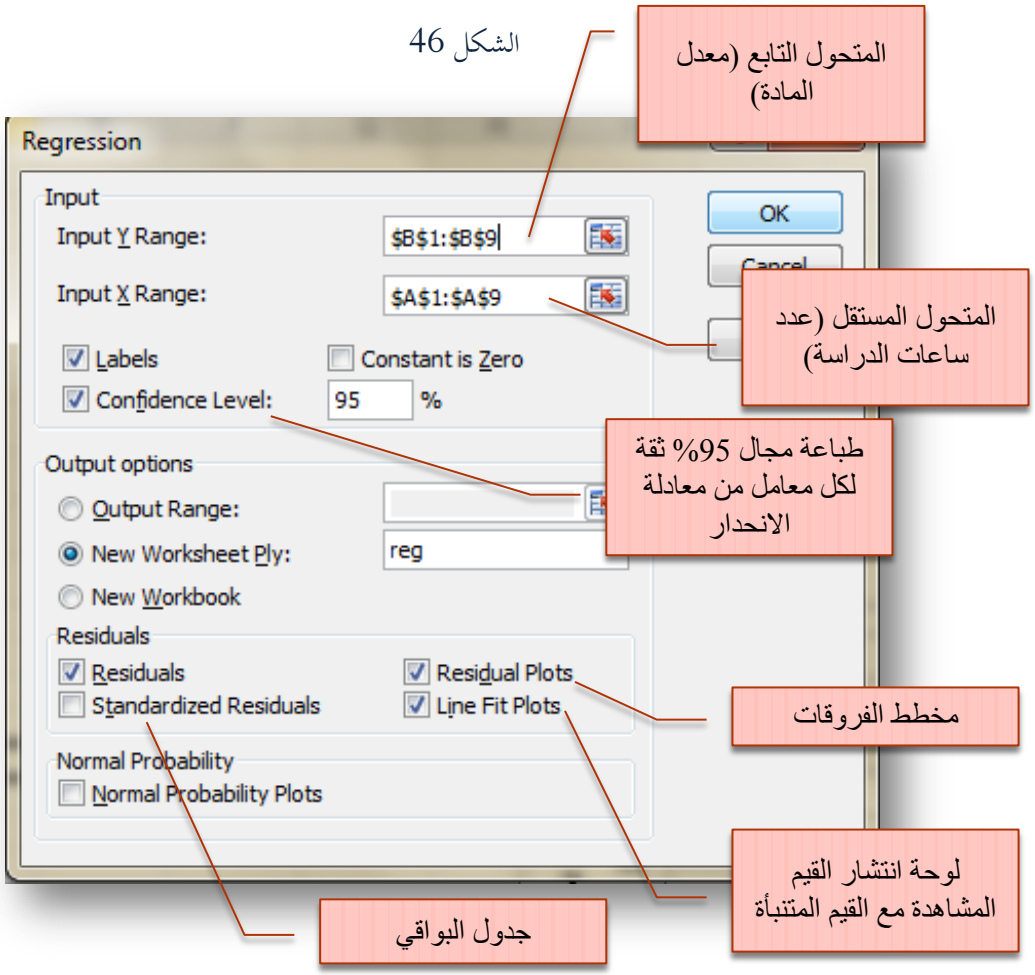
أدخل البيانات في صفحة إكسل كما يوضح الشكل 44 الآتي:

الشكل 45

	A	B
1	عدد ساعات الدراسة	معدل المادة
2	20	64
3	16	61
4	34	84
5	23	70
6	27	88
7	32	92
8	18	72
9	22	77
10		

اختر الوظيفة Regression من الأداة Data Analysis فتظهر لدينا النافذة الآتية (الشكل 46):

الشكل 46



فيظهر لدينا الخرج الآتي (الشكل 47 أ و ب):

الشكل 47 – أ

	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	*Regression Statistics*								
4	Multiple R	0.862108943							
5	R Square	0.74323183							
6	Adjusted R Square	0.700437135							
7	Standard Error	6.157605036							
8	Observations	8							
9									
10	ANOVA								
11		*df*	*SS*	*MS*	*F*	*Significance F*			
12	Regression	1	658.5034014	658.5034	17.36738	0.005895457			
13	Residual	6	227.4965986	37.9161					
14	Total	7	886						
15									
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
17	Intercept	40.08163265	8.889550951	4.508848	0.004065	18.32968508	61.83358023	18.32968508	61.83358023
18	عدد ساعات الدراسة	1.496598639	0.359118818	4.167419	0.005895	0.617866547	2.375330732	0.617866547	2.375330732
19									

تفسير المخرجات:

الجدول الأول: إحصاءات الانحدار Regression Statistics:

- Multiple R: وهو معامل الارتباط المتعدد بين المتحول التابع من جهة، وبين المتحولات المستقلة من جهة أخرى، وعندما يكون لدينا متحول مستقل وحيد، ينطبق هذا المعامل على معامل الارتباط البسيط بيرسون (الذي استعرضناه سابق في الفقرة 1.4)، وهو هنا يساوي 86% والذي يدل على وجود علاقة قوية طردية بين عدد ساعات الدراسة ومعدل المادة.
- R Square: وهو معامل التحديد، ويساوي مربع معامل الارتباط السابق. وتفيد هذه القيمة في تقيم جودة موافقة النموذج للبيانات. أي أنه يجيبنا عن السؤال: ما مدى دقة التنبؤ باستخدام النموذج المقدر؟ وفي مثالنا يساوي 74% ويقال هنا بأن النموذج يفسر سلوك 74% من البيانات، بمعنى أننا إذا أردنا التنبؤ اعتمادا على النموذج فإن النتائج ستكون دقيقة باحتمال 74%. لذا من الواضح أنه كلما ارتفعت قيمة معامل التحديد، دل ذلك على دقة أكبر وخطأ أقل في التنبؤ.
- Adjusted R Square: وهو معامل التحديد المعدل ويكافئ معامل التحديد السابق في استخدامه، لكن تم تعديله ليناسب حالة الانحدار الخطي المتعدد. وهو في هذه الحالة (أي في حالة وجود متحول مستقل وحيد، فإنه عادة لا يستخدم).
- : وهو الخطأ المعياري للتنبؤ، أي عند التنبؤ باستخدام هذا النموذج فإننا معرضون لخطأ تنبؤ مقداره 6.16 درجة.
- : وهو بالطبع عدد المشاهدات.

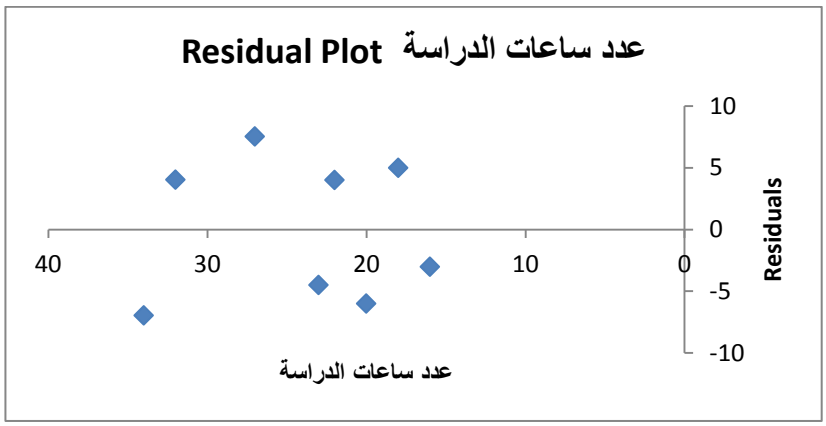
- **الجدول الثاني: تحليل التباين للانحدار ANOVA**: وهذا الجدول يختبر الفرضية الابتدائية القائلة بأن النموذج **لا** يصلح للتنبؤ، مقابل الفرضية البديلة بأن النموذج يصلح للتنبؤ. من الواضح أن معنوية هذا الاختبار (Significant F) هي 0.0059 وهي أصغر من مستوى الدلالة 0.05، وبالتالي فإننا نرفض الفرضية الابتدائية ونقبل بالبديلة القائلة بأن النموذج يصلح للتنبؤ (النموذج معنوي) عند مستوى دلالة 0.05
- **الجدول الثالث: جدول المعاملات:**
  - Coefficients: وهي معاملات النموذج، التقاطع هو 40.08، والميل هو 1.50، أي أن المعادلة الناتجة شكلها: Average = 40.08 + 1.5 (Hrs)
  - Standard Error: وهو الخطأ المعياري للمعامل المقدر، بمعنى أن التقاطع المقدر قيمته تساوي 40.08 بخطأ معياري 8.89، والميل المقدر قيمته 1.50 بخطأ معياري 0.36.
  - t Stat: وهو إحصاء اختبار الفرضية الابتدائية القائلة بأن قيمة المعامل يساوي الصفر مقابل الفرضية البديلة القائلة بأن المعامل لا يساوي الصفر.
  - P-value: وهي قيم معنوية كل اختبار فرضية، وهو بالنسبة للفرضية الأولى (للتقاطع) يساوي 0.004 وبالتالي فإننا نرفض الفرضية الابتائية القائلة بأن التقاطع يساوي الصفر، ونقبل البديلة القائلة بأن التقاطع لا يساوي الصفر عند مستوى دلالة 0.05 (أي أن التقاطع معنوي)، وبالنسبة للفرضية الثانية (المتعلقة بالميل) فمعنويته تساوي 0.006 وبالتالي فإننا نرفض الفرضية الابتائية القائلة بأن الميل يساوي الصفر، ونقبل البديلة القائلة بأن الميل لا يساوي الصفر عند مستوى دلالة 0.05 (أي أن الميل معنوي أيضا).
  - Lower 95% & Upper 95%: وهي حدود مجال 95% ثقة لكل معامل. على سبيل المثال يقع التقاطع الحقيقي في المجال [18.33, 61.83] باحتمال 95%. ونلاحظ أنها تتكرر مرتين، وهذه من (نهفات) إكسل!
- **الجدول الرابع: الفروقات (أو الرواسب) Residuals**: يقصد بالفروقات الفرق بين قيمة المتحول التابع المشاهد والقيمة المتنبأة له. على سبيل المثال الطالب الأول درس 20 ساعة وحصل في المادة على 64 درجة (الشكل 45)، تسمى الدرجة 64 بالقيمة المشاهدة للمتحول التابع، للحصول على القيمة المتنبأة المقابلة لهذا الطالب نستخدم معادلة الانحدار الناتجة Average = 40.08 + 1.5 (Hrs) ونعوض في المتحول المستقل القيمة Hrs = 20، فتصبح القيمة المتنبأة للمعدل: Average | $_{Hrs=20}$ = 40.08 + 1.5 (20) = 70.08. بمعنى أننا نتنبأ للطالب الذي يدرس 20 ساعة أنه سيحصل على 70.08 درجة، لكن القيمة الفعلية (المشاهدة) هي 64 درجة، أي نحن هنا أما خطأ مقداره – = 70.08 – 64 6.02 وهو قيمة الفرق بالنسبة للطالب الأول. ونلاحظ أنه كلما صغر هذا الفرق دل ذلك على نموذج أفضل (أي اقتراب القيمة المتنبأة من المشاهدة).

الشكل 47 – ب

	A	B	C
22	RESIDUAL OUTPUT		
23			
24	Observation	Predicted معدل المادة	Residuals
25	1	70.01360544	-6.013605442
26	2	64.02721088	-3.027210884
27	3	90.96598639	-6.965986395
28	4	74.50340136	-4.503401361
29	5	80.48979592	7.510204082
30	6	87.97278912	4.027210884
31	7	67.02040816	4.979591837
32	8	73.00680272	3.993197279
33			

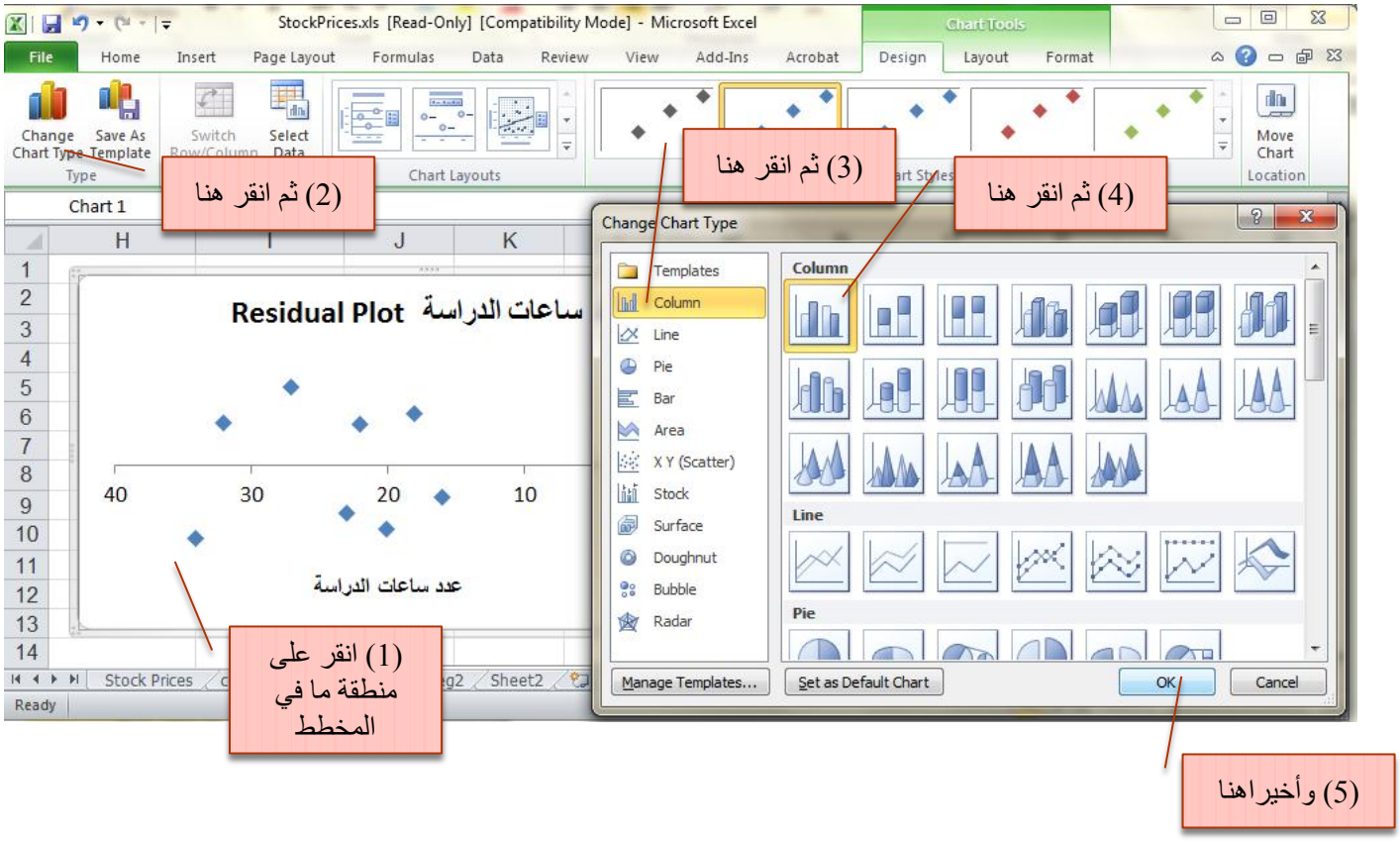
كما نلاحظ وجود مخطط بياني مرفق مع المخرجات وهو مخطط الفروقات (الشكل 48):

الشكل 48



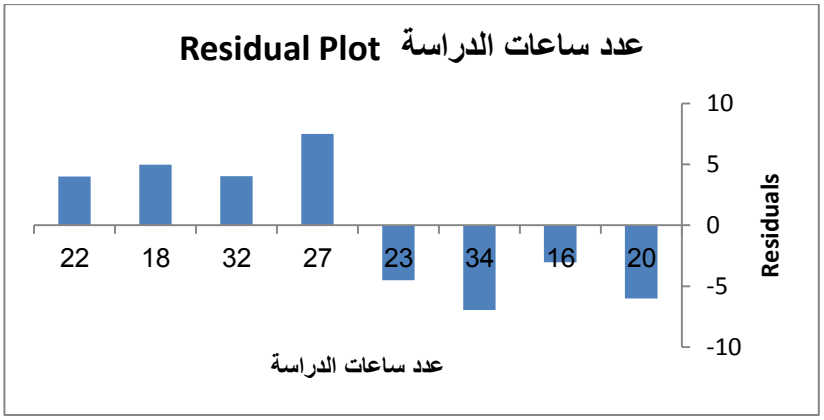
تمثل النقط أماكن توضع قيم الفروق (الموضحة في جدول الشكل 47)، لكن عادة ما يتم تمثيل الفروقات بالأعمدة، لجعل الشكل أعمدة اتبع الخطوات الموضحة في الشكل 49:

الشكل 49



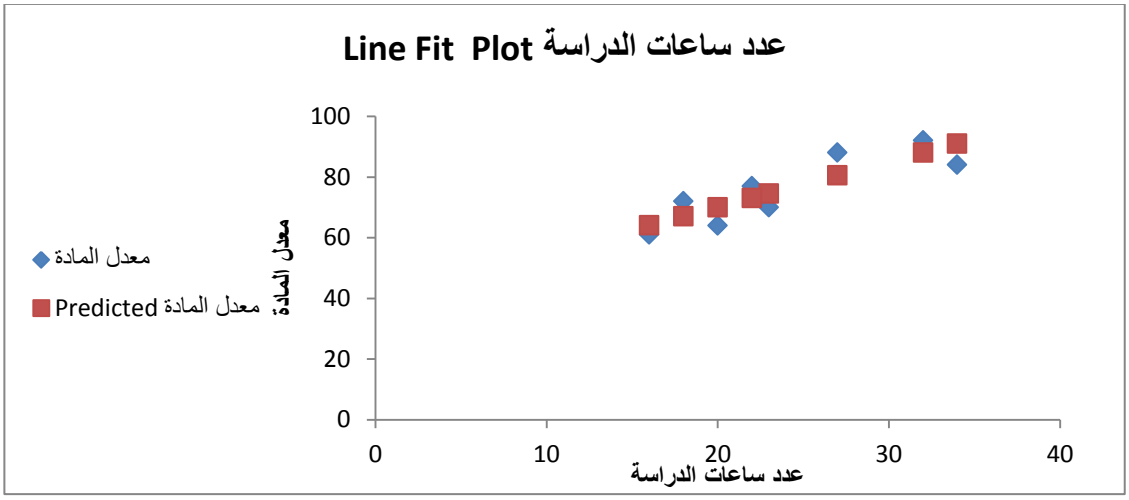
فيصبح المخطط كما يوضحه الشكل 50:

الشكل 50



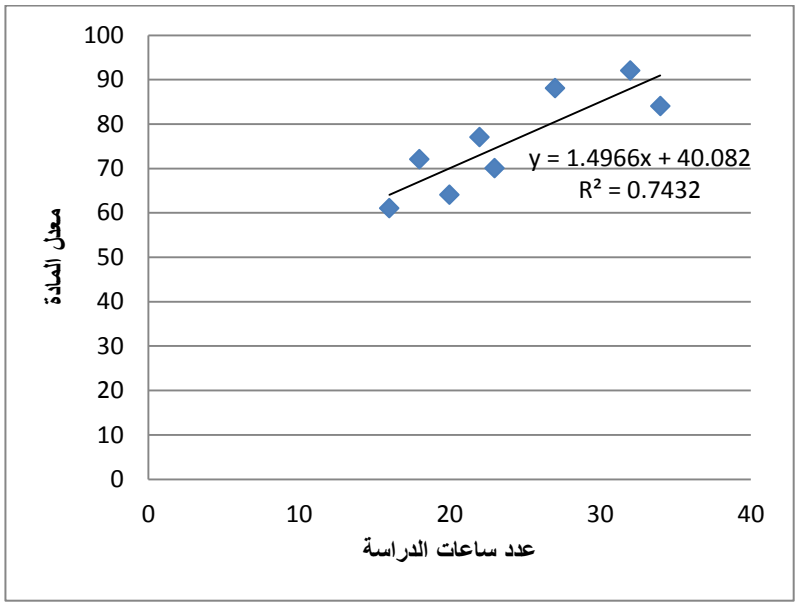
كما يظهر مع المخرجات لوحة الانتشار للقيم المشاهدة مع القيمة المتنبأة (الشكل 51):

الشكل 51



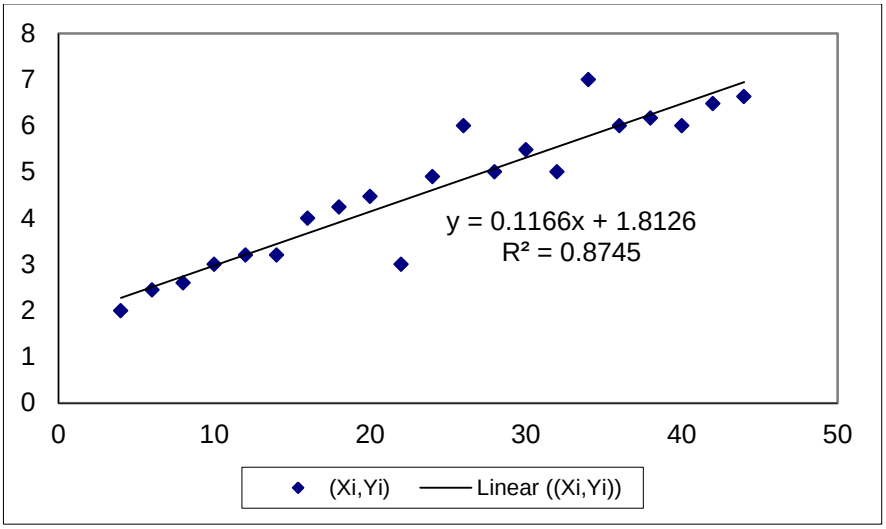
أما معادلة الانحدار الخطي البسيط معادلة المستقيم ((Hrs) 1.5 + 40.08 = Average)، فموضحة بالشكل 52 الآتي:

الشكل 52

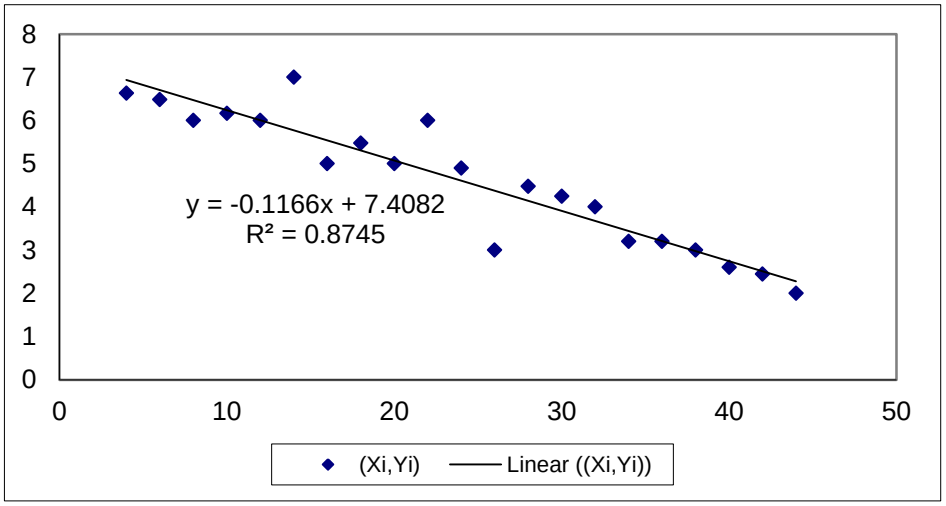


وكمثال آخر لنأخذ الأشكال الآتية 53 أ و ب:

الشكل 53 – أ



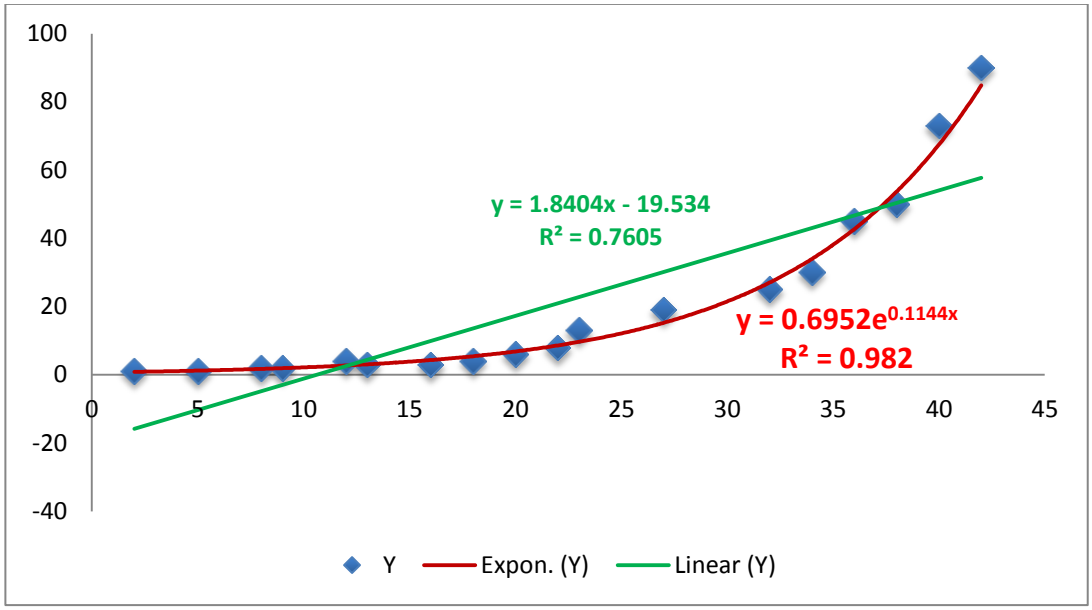
الشكل 53– ب



لاحظ أن ميل معادلة الشكل 53 – أ موجب والذي يعني أنه مع زيادة قيمة X تزداد قيمة Y، بينما ميل معادلة الشكل 53– ب سالب، والذي يعني أنه مع زيادة قيمة X تنقص قيمة Y، لكن من الملاحظ أن معامل التحديد في الشكل 53 (أ وب) يساوي 87% هو أكبر معامل التحديد في الشكل 52 والذي يساوي 74%، وذلك بسبب أن النقط في الشكل 53 أكثر تمركزا حول الخط المستقيم مقارنة مع تمركز النقط في الشكل 52.

لكن لننظر إلى الشكل 54 الآتي:

الشكل 54



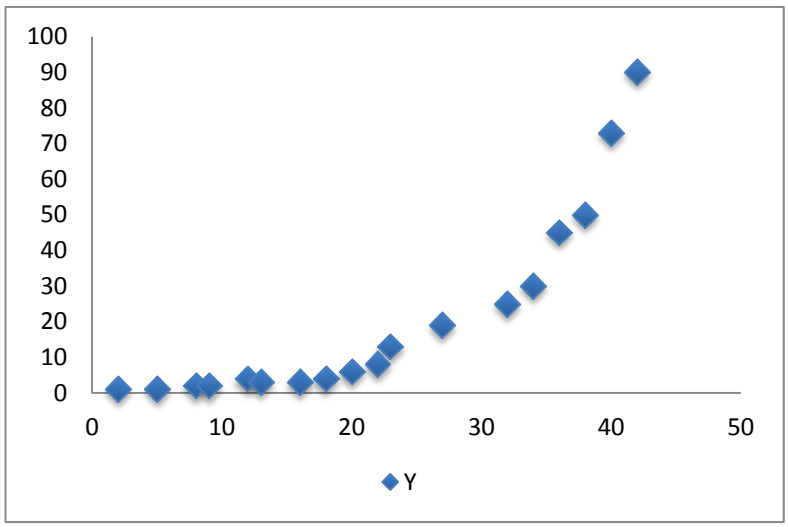
بمقارنة الخط المستقيم (معادلة الانحدار الخطي البسيط) مع الخط المنحني الصاعد (معادلة الانحدار الأسي البسيط)، وبالنظر أيضا إلى معامل التحديد، نجد أن معامل التحديد للنموذج الأسي (98%) أكبر من معامل التحديد للنموذج الخطي (76%)، والذي يدل على أن معادلة النموذج الأسي أكثر توافقا مع البيانات من معادلة النموذج الخطي.

لذا عموما، توجد نماذج مختلفة عن النموذج الخطي قد تكون أكثر توافقا من النموذج الخطي. يمكن توقع النموذج الأكثر توافق بالنظرة السريعة إلى لوحة الانتشار. ويمكن رسم وإيجاد معادلة الانحدار بأشكال مختلفة مباشرة في إكسل من خلال الخطوات:

- ارسم لوحة الانتشار للمتحولين Scatter Plot المستقل والتابع (سنعرض في الفصل الخامس كيفية رسم لوحة الانتشار).

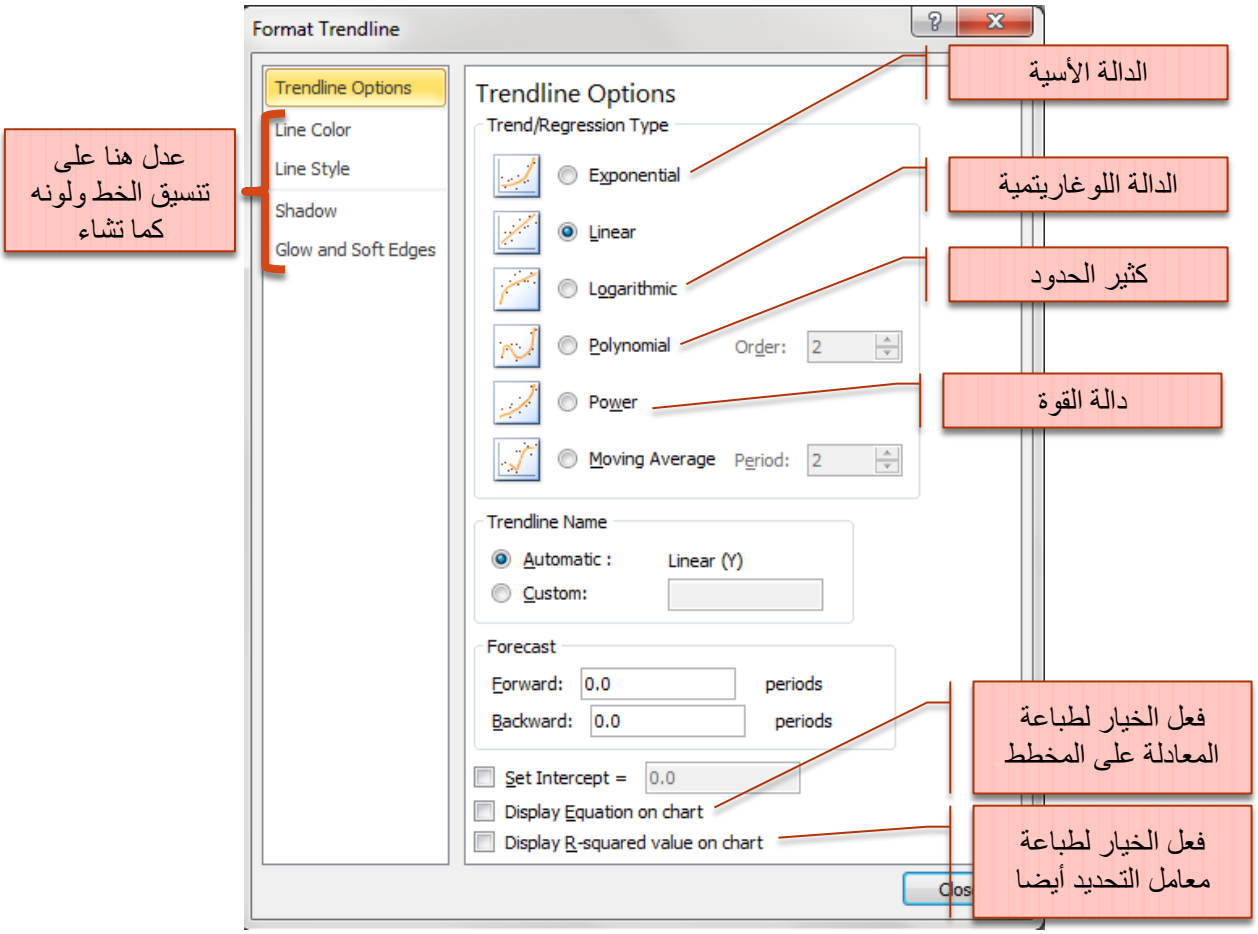
ولنفرض أن لوحة الانتشار كان لها الشكل 55 الآتي:

الشكل 55

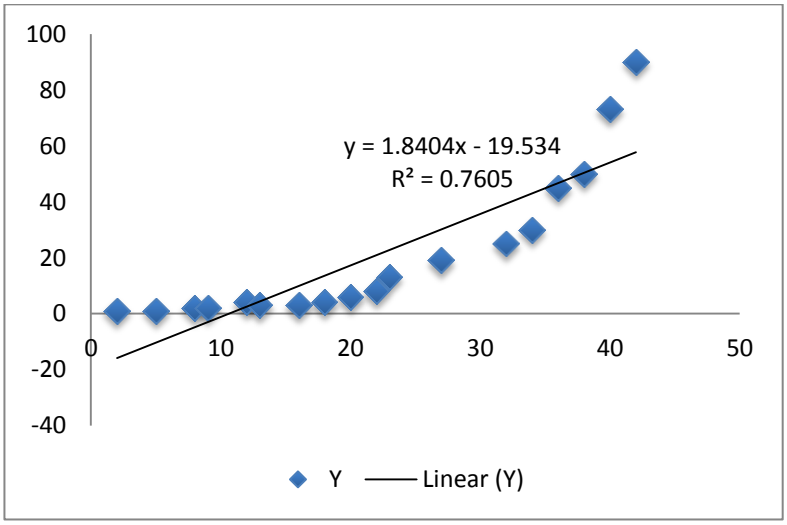


انقر بزر الفأرة الأيمن فوق إحدى النقط لتظهر قائمة، اختر منها (إضافة خط اتجاه – Add trend line). ثم تابع كما يوضح الشكل 56 الآتي:

الشكل 56



ليظهر الخطي المستقيم مع المعادلة ومعامل التحديد كما يوضح الشكل 57 الآتي:



استعرضنا حتى الآن معادلة الانحدار الخطي البسيط والذي يفترض وجود متحول مستقل واحد مؤثر في المتحول التابع، سنضيف في الفقرة الآتية أكثر من متحول واحد إلى معادلة الانحدار لنحصل على ما يسمى معادلة الانحدار الخطي المتعدد.

3.4 الانحدار الخطي المتعدد:

لتكن البيانات التالية لعينة عشوائية مسحوبة من موظفي قسم التقنيات الإلكترونية لشركة اتصالات:

رقم الوحدة	الراتب السنوي Salary	عدد سنوات الخبرة Exp	عدد سنوات التعليم الجامعي Edu	الجنس Gender
1	54900	5.5	4	F
2	60500	9	4	M
3	58900	4	5	F
4	59000	8	4	M
5	57500	9.5	5	M
6	55500	3	4	F
7	56000	7	3	F
8	52700	1.5	4.5	F
9	65000	8.5	5	M
10	60000	7.5	6	F
11	56000	9.5	2	M
12	53600	6	2	F
13	55000	2.5	4	M
14	52500	1.5	4.5	M

(1) أوجد معادلة الانحدار الخطي المتعدد للمتحولات المستقلة الثلاثة على المتحول التابع (الراتب السنوي).

(2) أعد تقدير معادلة الانحدار المعنوية

(3) بفرض أنه يوجد موظفة خريجة صيدلة في هذا القسم وعدد سنوات خبرتها عشر سنوات، كم تتوقع راتبها السنوي؟

(4) هل يمكننا إضافة عامل العمر أيضا إلى جانب المتحولات المستقلة البقية؟

رتب أولا البيانات في صفحة إكسل كما يوضح الشكل 58 الآتي:

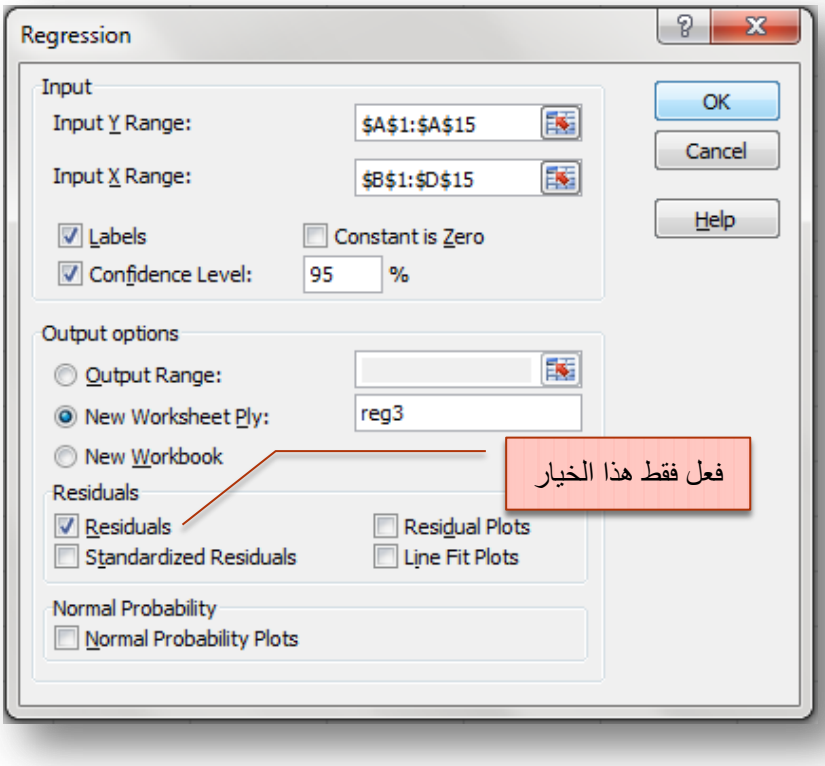
الشكل 58

	A	B	C	D	E
1	الراتب السنوي Salary	عدد سنوات الخبرة Exp	عدد سنوات التعليم الجامعي Edu	الجنس Gender	
2	54900	5.5	4	0	
3	60500	9	4	1	
4	58900	4	5	0	
5	59000	8	4	1	
6	57500	9.5	5	1	
7	55500	3	4	0	
8	56000	7	3	0	
9	52700	1.5	4.5	0	
10	65000	8.5	5	1	
11	60000	7.5	6	0	
12	56000	9.5	2	1	
13	53600	6	2	0	
14	55000	2.5	4	1	
15	52500	1.5	4.5	1	
16					

لاحظ كيف استبدلنا تسميات الذكور (M) والإناث (F) بالأرقام 1 و0 على الترتيب (تشفير البيانات الإسمية) لأن تحليل الانحدار ببساطة لا يتعامل إلا مع الأرقام.

اختر الآن الوظيفة Regression من الأداة Data Analysis فتظهر لدينا النافذة الآتية (الشكل 59):

الشكل 59



لتظهر المخرجات الآتية (الشكل 60 أ و ب):

الشكل 60 – أ

	A	B	C	D	E	F	G
1	SUMMARY OUTPUT						
2							
3	*Regression Statistics*						
4	Multiple R	0.8215905					
5	R Square	0.6750109					
6	Adjusted R Square	0.5775142					
7	Standard Error	2251.7158					
8	Observations	14					
9							
10	ANOVA						
11		*df*	*SS*	*MS*	*F*	*Significance F*	
12	Regression	3	105309901.3	35103300	6.923422	0.008376182	
13	Residual	10	50702241.6	5070224			
14	Total	13	156012142.9				
15							
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
17	Intercept	45495.329	2809.771399	16.19183	1.67E-08	39234.76856	51755.89019
18	Exp عدد سنوات الخبرة	801.5711	228.4684282	3.508455	0.005646	292.5117211	1310.630484
19	Edu عدد سنوات التعليم الجامعي	1595.7365	560.6439792	2.846256	0.017361	346.5439059	2844.92917
20	Gender الجنس	382.57208	1287.410877	0.297164	0.772423	-2485.958112	3251.102273
21							

تفسير المخرجات:

- **الجدول الأول:**
  - معامل الارتباط المتعدد (Multiple R) يساوي 0.82 وبالتالي العلاقة بين المتحول التابع من جهة (المرتب السنوي)، وبين المتحولات المستقلة من جهة أخرى علاقة قوية طردية، بمعنى أنه مع ازدياد قيم المتحولات المستقلة المعنوية يزداد المرتب السنوي.
  - معامل التحديد (R Square) يساوي 0.68 وبالتالي فإن النموذج يفسر 68% من البيانات. بمعنى أننا إذا تنبأنا باستخدام هذا النموذج سنحصل على نتائج دقيقة باحتمال 68%.
  - معامل التحديد المعدل (Adjusted R Square) يساوي 0.58، وهو يفيد في اختيار النموذج الأكثر معنوية، إذ أن هذا المقياس يزيد في إحدى الحالتين: إضافة متحول مستقل معنوي إلى المعادلة، أو حذف متحول مستقل غير معنوي من المعادلة كما سنرى لاحقا عندما سنحذف متحول الجنس غير المعنوي.
  - الخطأ المعياري للتقدير (Standard Error) يساوي $2,251.7
- **جدول تحيل التباين:** يتضح من قيمة المعنوية (Significant F = 0.0084) أن النموذج معنوي، بمعنى أن النموذج يصلح للتنبؤ.
- **جدول المعاملات:**

- معامل التقاطع: يساوي 45,495.329 وهو معنوي (1.67×10^{-8}=1.67E-08)
- معامل متحول عدد سنوات الخبرة: يساوي 801.57 وهو معنوي (0.0056)، ويلاحظ بأنه قيمته الموجبة تدل على العلاقة الطردية بين عدد سنوات الخبرة والراتب السنوي. بمعنى أنه مع زيادة عدد سنوات الخبرة يزداد المرتب السنوي.
- معامل متحول عدد سنوات التعليم الجامعي: يساوي 1,595.74 وهو معنوي (0.017)، ويلاحظ أيضا من قيمته الموجبة أن علاقته طردية مع الراتب السنوي. ونلاحظ أن معنويته أكبر من معنوية معامل عدد سنوات الخبرة، الذي يدل على أن عدد سنوات الخبرة أكثر تاثير في الراتب السنوي من عدد سنوات التعليم الجامعي.
- معامل متحول الجنس: يساوي 382.57 لكنه غير معنوي (0.772) بمعنى أنه لا يوجد تمييز بين مرتبات الذكور والإناث في هذه المؤسسة.

الشكل 60– ب

تؤخذ المعلومات من السطر الأول في الشكل 58 وتحسب القيمة المتوقعة لهذا الراتب مثلا بالشكل:
56,286.9 =
45,495.3+801.6(5.5)+1,595.7(4)+382.6(0)

23			
24	RESIDUAL OUTPUT		
25			
26	*Observation*	*Predicted الراتب السنوي Salary*	*Residuals*
27	1	56286.91659	-1386.916592
28	2	59474.98753	1025.012469
29	3	56680.29648	2219.703524
30	4	58673.41643	326.5835715
31	5	61471.50962	-3971.50962
32	6	54282.98884	1217.011165
33	7	55893.53671	106.463293
34	8	53878.50045	-1178.500451
35	9	60669.93852	4330.061482
36	10	61081.53187	-1081.531873
37	11	56684.30001	-684.3000058
38	12	53496.22907	103.7709336
39	13	54264.77536	735.224635
40	14	54261.07253	-1761.072532
41			

الفرق بين الراتب المتنبأ به باستخدام المعادلة (56,286.9) والراتب الفعلي للمشاهدة الأولى (السطر الأول في الشكل 58)

بالطبع كلما صغرت قيم الفروقات دل ذلك على أن النموذج أفضل للتنبؤ، والنتائجة أدق.

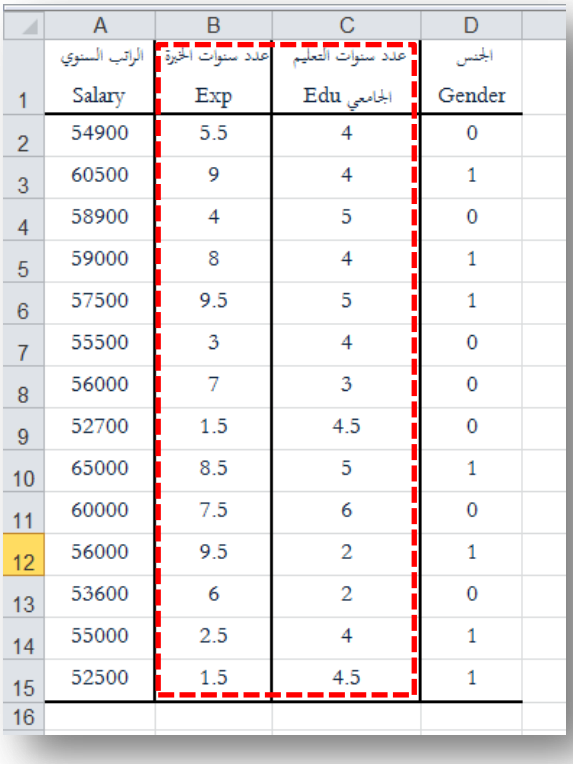
الخلاصة: وجد أن أكثر عامل مؤثر في الراتب السنوي هو الخبرة، ثم يأتي عدد سنوات التعليم، لكن لم يكن هنالك تأثير للجنس في الراتب السنوي. والمعادلة الناتجة هي:

Salary = 45,495.3 + 801.6 Experience + 1,595.7 Education + 382.6 Gender

لكن يوجد في هذه المعادلة متحول الجنس الذي يؤثر وجوده على نتائج التنبؤ، لكنه بالمقابل ليس معنويا. لذا لا بد من حذفه من المعادلة للحصول على نتائج تنبؤ أدق. أي لا بد من إعادة إيجاد معادلة دالة الانحدار لكن هذه المرة بدون إدخال متحول الجنس في المعادلة بمعنى

أن نطاق المتحولات المستقلة الجديد في الخلية (Input X Range) في نافذة مدخلات الانحدار (Regression) يوضحه الشكل 61 الآتي:

الشكل 61



	A	B	C	D
1	الراتب السنوي Salary	عدد سنوات الخبرة Exp	عدد سنوات التعليم الجامعي Edu	الجنس Gender
2	54900	5.5	4	0
3	60500	9	4	1
4	58900	4	5	0
5	59000	8	4	1
6	57500	9.5	5	1
7	55500	3	4	0
8	56000	7	3	0
9	52700	1.5	4.5	0
10	65000	8.5	5	1
11	60000	7.5	6	0
12	56000	9.5	2	1
13	53600	6	2	0
14	55000	2.5	4	1
15	52500	1.5	4.5	1
16				

أما نطاق المتحول التابع فيبقى على حاله، فتظهر لدينا المخرجات الموضحة في الشكل 62 والذي يقارن أيضا بين المخرجات قبل وبعد حذف متحول الجنس :

الشكل 62 – أ

قبل حذف متحول الجنس

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.821590487
R Square	0.675010928
Adjusted R Square	0.577514206
Standard Error	2251.715826
Observations	14

بعد حذف متحول الجنس

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.819842
R Square	0.672141
Adjusted R Square	0.61253
Standard Error	2156.385
Observations	14

نلاحظ أن قيم معاملات الارتباط المتعدد لم تتغير تقريبا، وكذلك معامل التحديد. والسبب في ذلك أن متحول الجنس لم يكن يساهم في الأصل في النموذج معامل معنوي، بمعنى أنه لا توجد أية علاقة بين متحول الجنس ومتحول الراتب التابع، لذا فإن حذفه أو إضافته لا تؤثر في علاقة الارتباط المتعدد، لكن بالنظر إلى قيمة معامل التحديد المعدل (Adjusted R Square) نلاحظ أن قيمته قد ازدادت

بعد حذف متحول الجنس المستقل من المعادلة وهذا دليل على أن متحول الجنس غير معنوي، إذا أن قيمة معامل التحديد المعدل تزداد عند حذف متحول مستقل غير معنوي، أو عند إضافة متحول مستقل معنوي.

الشكل 62 – ب

قبل حذف متحول الجنس

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	105309901	35103300	6.923	0.008376182
Residual	10	50702242	5070224		
Total	13	156012143			

بعد حذف متحول الجنس

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	104862168	52431084	11.28	0.002169095
Residual	11	51149975	4649998		
Total	13	156012143			

نلاحظ أن معنوية النموذج قد ازدادت بعد حذف متحول الجنس من 0.008 إلى 0.002 بمعنى أن النموذج الجديد أفضل للتنبؤ من النموذج السابق.

الشكل 62 – ج

قبل حذف متحول الجنس

	Coefficien	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	45495	2809.771399	16.2	2E-08	39234.77	51755.89
عدد سنوات الخبرة Exp	801.6	228.4684282	3.51	0.006	292.5117	1310.6305
عدد سنوات التعليم الجامعي	1596	560.6439792	2.85	0.017	346.5439	2844.9292
الجنس Gender	382.6	1287.410877	0.3	0.772	-2485.958	3251.1023

بعد حذف متحول الجنس

	Coefficien	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
ntercept	45511	2690.32008	16.92	3E-09	39589.982	51432.69
عدد سنوات الخبرة Exp	825.67	204.5508872	4.036	0.002	375.45454	1275.881
عدد سنوات التعليم الجامعي	1603.7	536.2945512	2.99	0.012	423.32265	2784.075

من الواضح أن المتحولات المستقلة المعنوية قد حافظت على معنويتها.

الشكل 63 – د

قبل حذف متحول الجنس

بعد حذف متحول الجنس

RESIDUAL OUTPUT

erva	Predicted الراتب السنوي Salary	Residuals
1	56286.91659	-1386.917
2	59474.98753	1025.012
3	56680.29648	2219.704
4	58673.41643	326.5836
5	61471.50962	-3971.51
6	54282.98884	1217.011
7	55893.53671	106.4633
8	53878.50045	-1178.5
9	60669.93852	4330.061
10	61081.53187	-1081.532
11	56684.30001	-684.3
12	53496.22907	103.7709
13	54264.77536	735.2246
14	54261.07253	-1761.073

RESIDUAL OUTPUT

erva	Predicted الراتب السنوي Salary	Residuals
1	56467.30664	-1567.31
2	59357.14467	1142.855
3	56832.50362	2067.496
4	58531.47666	468.5233
5	61373.67767	-3873.68
6	54403.13661	1096.863
7	56102.10966	-102.11
8	53966.4841	-1266.48
9	60548.00966	4451.99
10	61326.04065	-1326.04
11	56562.58069	-562.581
12	53672.74265	-72.7427
13	53990.30261	1009.697
14	53966.4841	-1466.48

من الواضح أن قيم الفروقات قد صغرت بعد حذف متحول الجنس، الأمر الذي يدل على حصولنا على نموذج أفضل وأدق للتنبؤ.

في المحصلة، فإن النموذج المعنوي النهائي:

$$Salary = 45,511.3 + 825.6\ Experience + 1,603.7\ Education$$

على سبيل المثال، إذا علمت أن موظفا ما في الشركة خبرته في الشركة 10 سنوات، وهو مهندس معلوماتية (أي عدد سنوات تعليمه الجامعي 5 سنوات)، فكم تتوقع دخله السنوي في هذه الشركة؟ الإجابة على هذا السؤال من خلال المعادلة السابقة:

$$Salary = 45,511.3 + 825.6\ (10) + 1,603.7\ (5) = 61,785.8\$$$

5. الرسم البياني في إكسل:

يمتلك إكسل إمكانيات لا مثيل لها في المخططات البيانية بمختلف أنواعها، فضلا عن جمالية المخططات التي يقدمها وإمكانية التحكم بكل خصائص وتنسيقات المخطط، يوضح الشكل 64 الآتي كيفية الوصل إلى أدوات الرسم البياني:

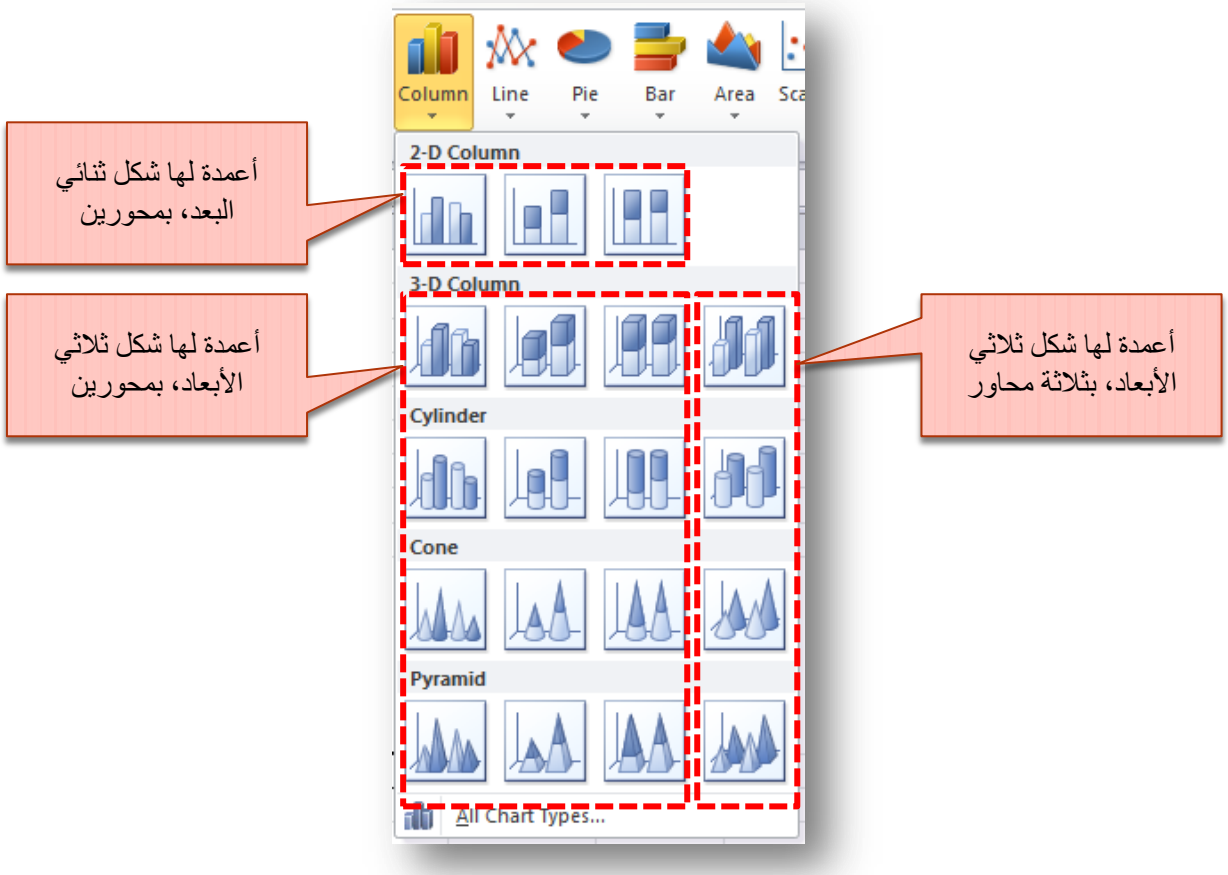
الشكل 64

كما أن لهذه المخططات ميزة غير موجودة في أي برنامج إحصائي آخر، وهي ديناميكية المخططيات، بمعنى أن المخططات تتغير مباشرة عند تغيير أي قيمة للبيانات التي تتبع لها المخططات كما سنرى لاحقا.

1.5 مخططات الأعمدة Bar Chart:

يوضح الشكل 65 الآتي ما تتضمنه مخططات الأعمدة:

الشكل 65



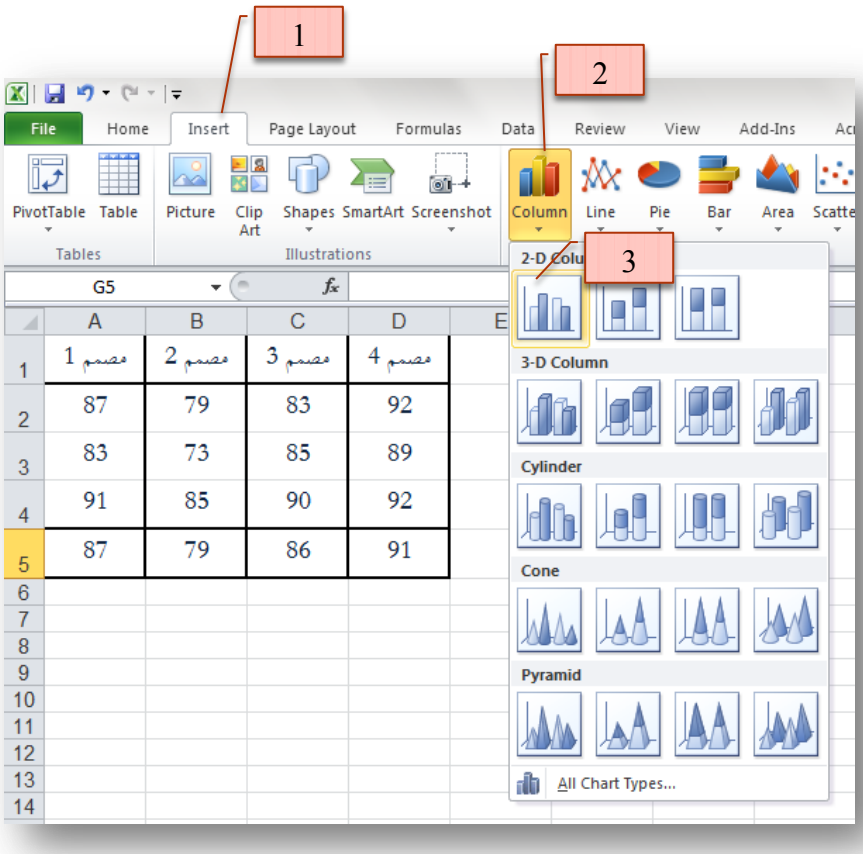
لنأخذ أولا بيانات مقارنة أداء المصممين (الفقرة 1.3) ولنكتبها مع متوسط تقييم كما هو موضح في الشكل 66 الآتي:

الشكل 66

	A	B	C	D
1	مصمم 1	مصمم 2	مصمم 3	مصمم 4
2	87	79	83	92
3	83	73	85	89
4	91	85	90	92
5	87	79	86	91

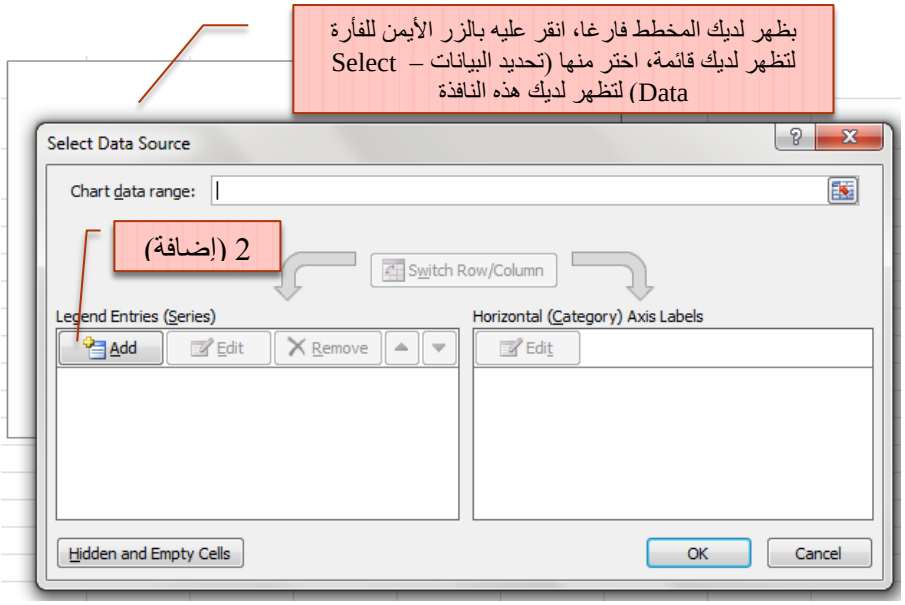
ولنفرض أننا نريد مخطط الأعمدة للمتوسطات، عندها نتبع الخطوات الموضحة في الأشكال 67 الآتية:

الشكل 67 – أ



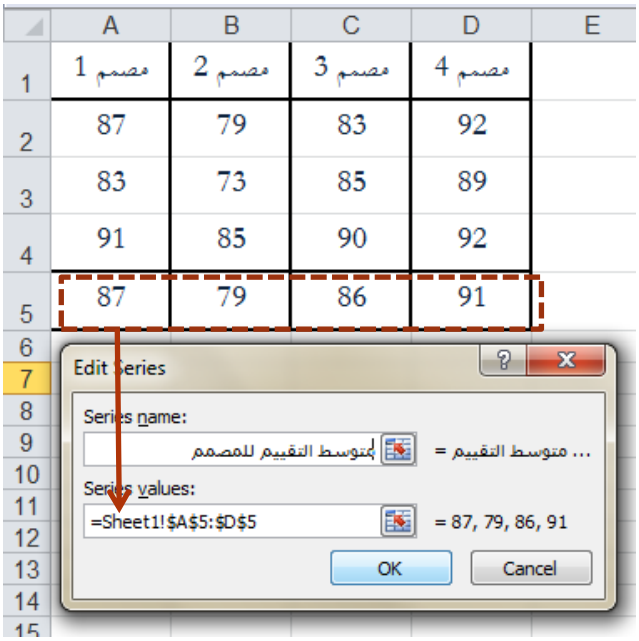
قم أولا بإدراج المخطط على صفحة إكسل

الشكل 67 – ب



نافذة تحرير بيانات المخطط، بالنقر على (إضافة – Add) تظهر لدينا النافذة الموضحة في الشكل 67 – ج الآتي:

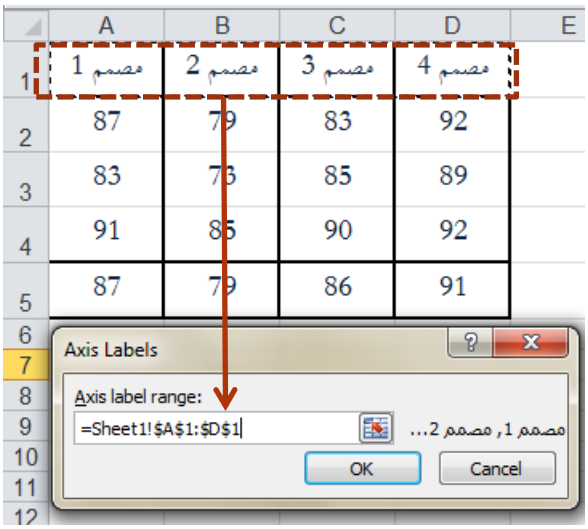
الشكل 67 – ج



نملأ الحقول كما هو موضح في الشكل المجاور ثم بالنقر على موافق (OK) نعود للنافذة في الشكل 67 – ب ننقر في هذه النافذة وضمن الإطار المسماة (Horizontal (Category) Axis Lables – المحور الأفقي) على (**تحرير**) **(Edit –**

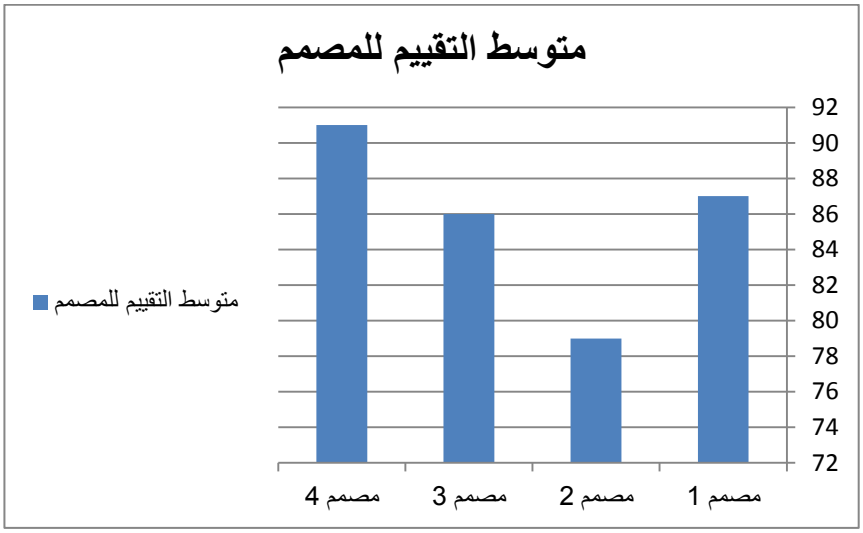
لتظهر النافذة الموضحة في الشكل 67 – د الآتي:

الشكل 67 – د

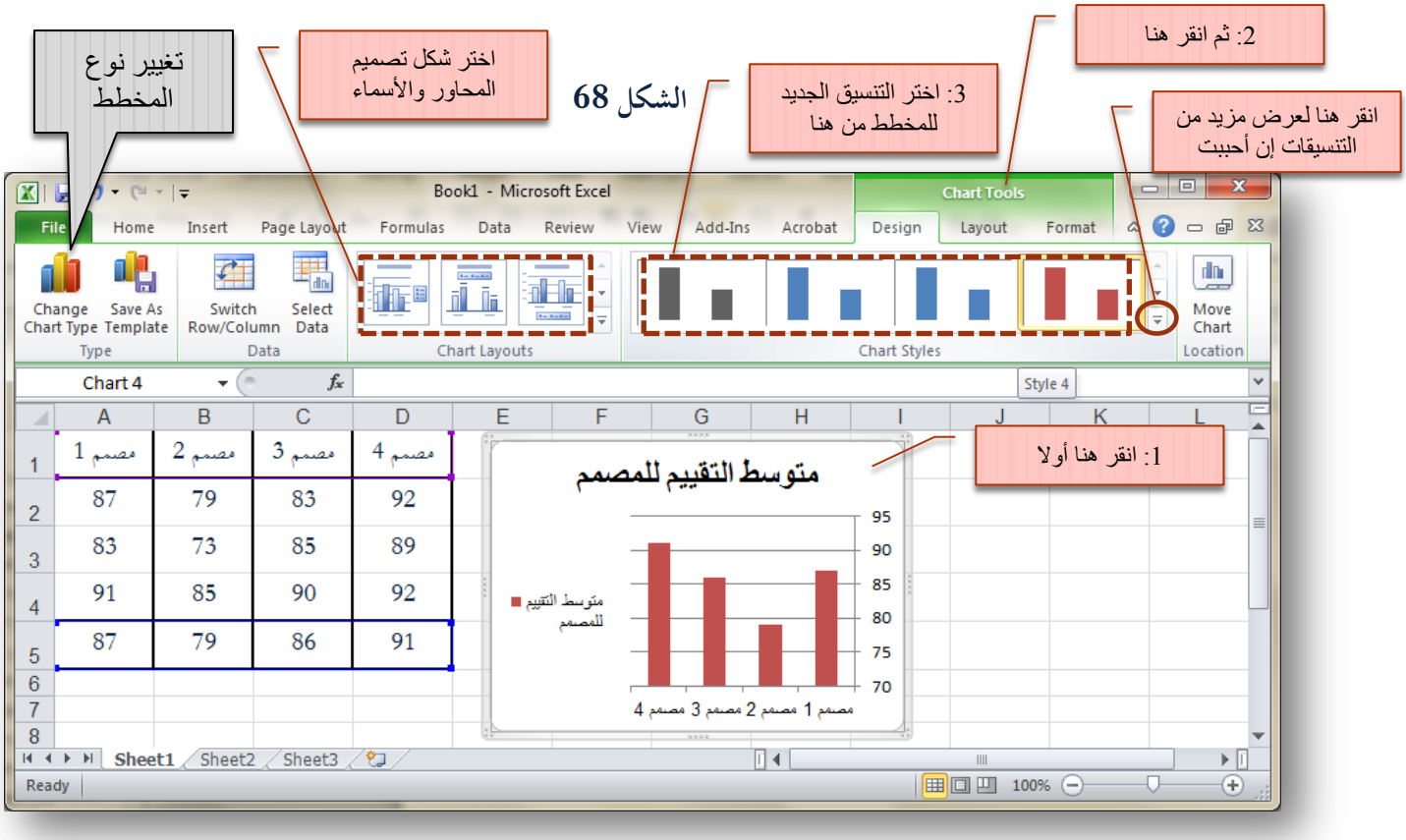


نملأ الحقل كما هو موضح في الشكل المجاور ثم بالنقر على نعود للنافذة في الشكل 67 – ب ننقر في هذه النافذة أيضا على موافق (OK) ليظهر المخطط النهائي كما هو موضح في الشكل 67 – هـ الآتي:

الشكل 67– هـ

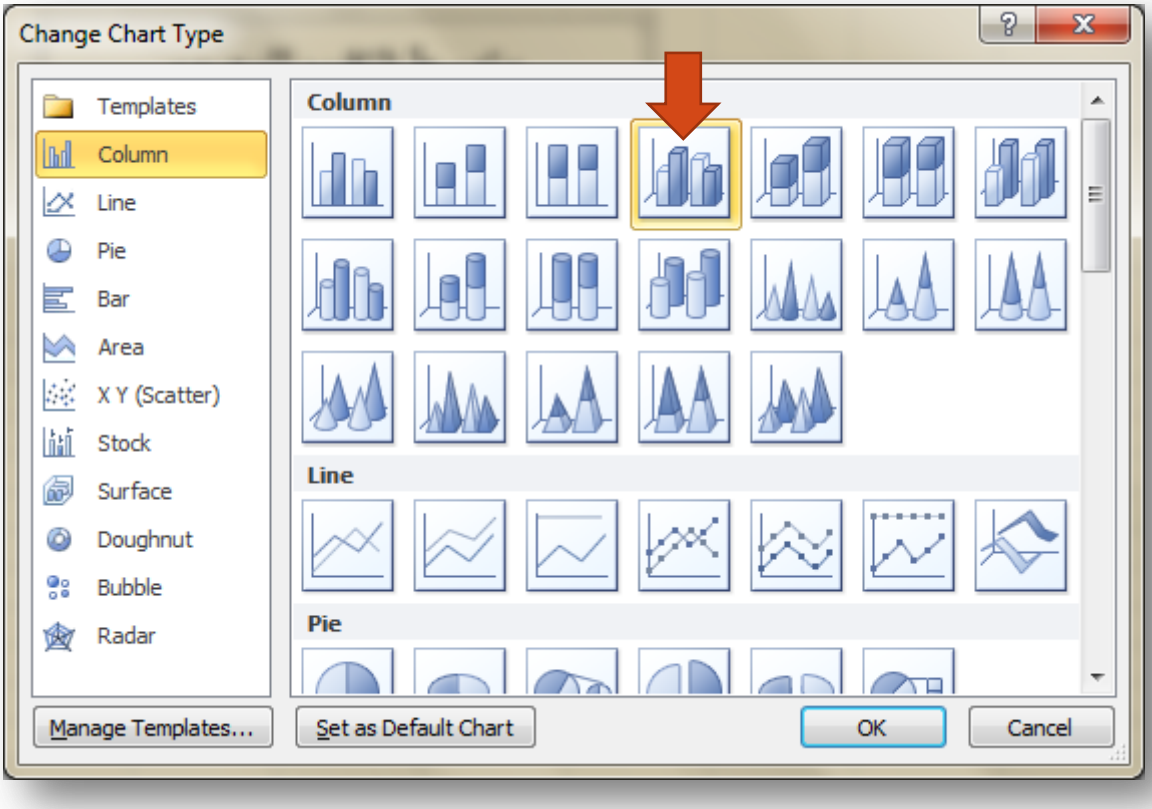


بالطبع يمكنك تغيير ألوان وتنسيقات هذا المخطط كما هو موضح في الشكل 68 الآتي، سأترك لك تجريب هذه الخيارات:



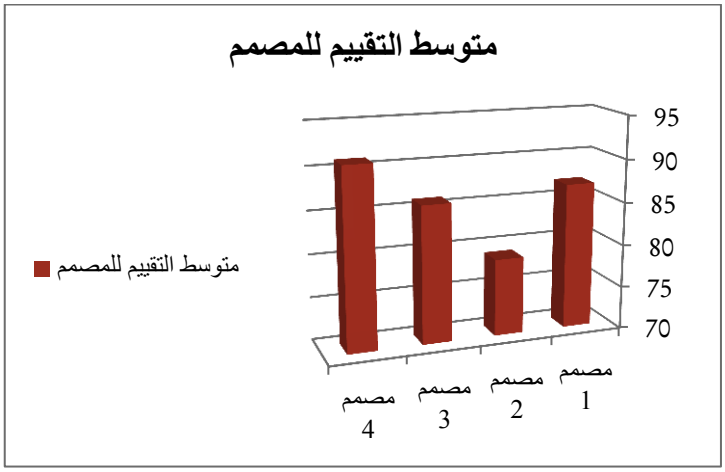
كما تستطيع جعل هذه الأعمدة ثلاثية الأبعاد من خلال أيقونة (تغيير نوع المخطط – Change Chart Type) كما هو موضح في يسار الشكل 68 السابق لتظهر النافذة الآتية (الشكل 69):

الشكل 69



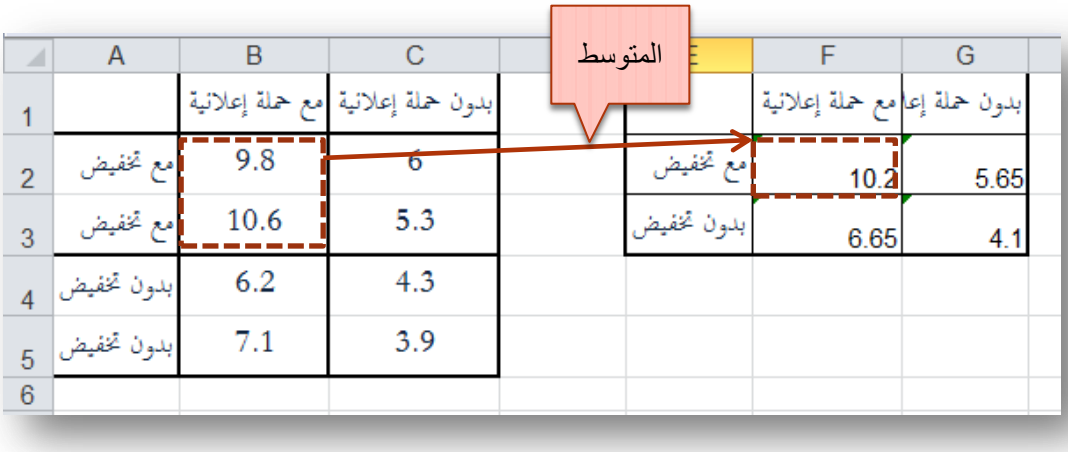
ليصبح المخطط كما هو موضح في الشكل 70 الآتي:

الشكل 70



لنأخذ حالة أخرى وهي كون لدينا بيانات ثنائية البعد (أي تحتاج إلى بعدين)، ولتكن بيانات دراسة تأثير كل من المبيعات والتخفيضات على المبيعات (الفقرة 2.3) والموضحة في الشكل 25، أضف إلى هذه البيانات متوسط مبيعات كل تصنيف كما يوضح الشكل 71 الآتي:

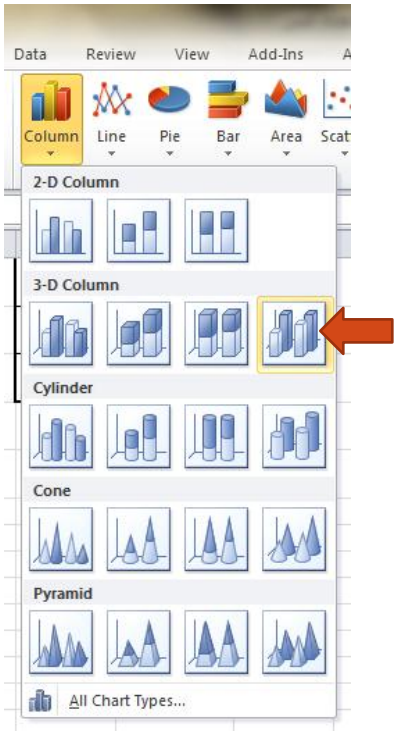
الشكل 71



	A	B	C	E	F	G
1		مع حملة إعلانية	بدون حملة إعلانية		مع حملة إعلانية	بدون حملة إع
2	مع تخفيض	9.8	6	مع تخفيض	10.2	5.65
3	مع تخفيض	10.6	5.3	بدون تخفيض	6.65	4.1
4	بدون تخفيض	6.2	4.3			
5	بدون تخفيض	7.1	3.9			
6						

الآن باختيار المخطط الموضح في الشكل 72 الآتي:

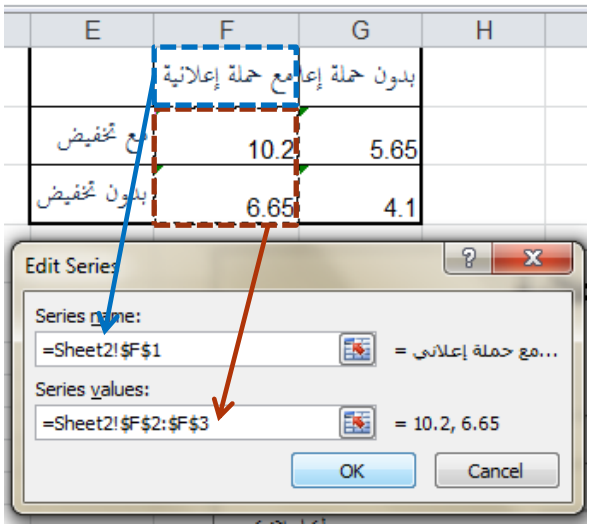
الشكل 72



فيظهر على صفحة إكسل مخطط فارغ، ننقر على هذا المخطط الفارغ بالزر الأيمن للفأرة، نختر من القائمة الظاهرة (تحديد البيانات – Select Data) لتظهر نافذة (تحديد مصدر البيانات – Select Data Source) تماما كما يوضحها الشكل 67 – ب السابق.

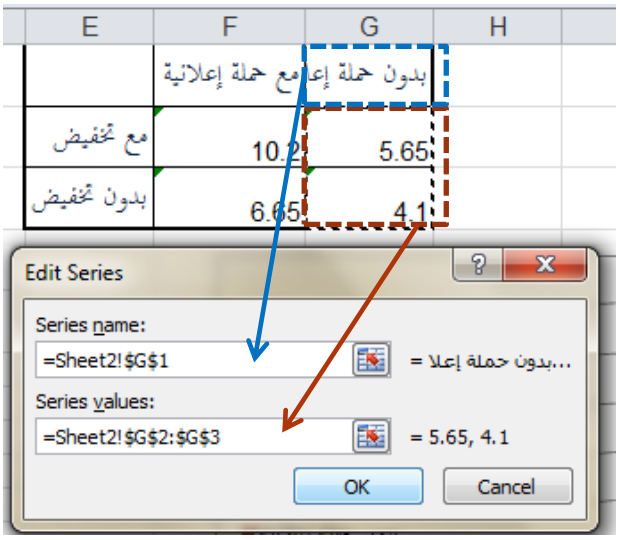
ننقر في هذه النافذة على (إضافة – Add) لتظهر النافذة الآتية (الشكل 73– أ):

الشكل 73 – أ



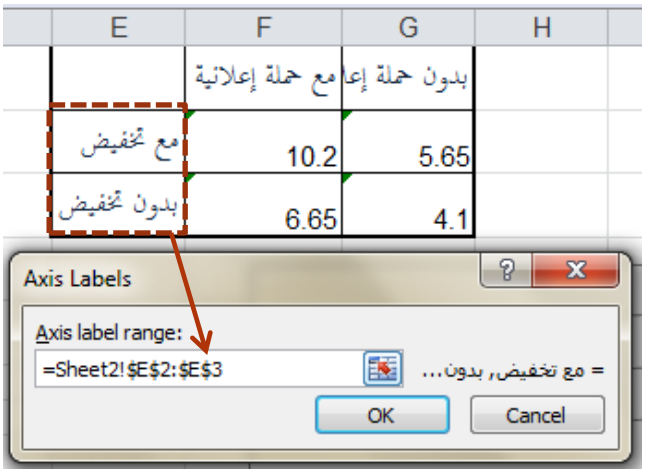
ننقر على (موافق– OK) لنعود للنافذة (تحديد مصدر البيانات – Select Data Source) والتي سنعيد فيها النقر على (إضافة – Add) لتظهر النافذة الآتية (الشكل 73– ب):

الشكل 73– ب



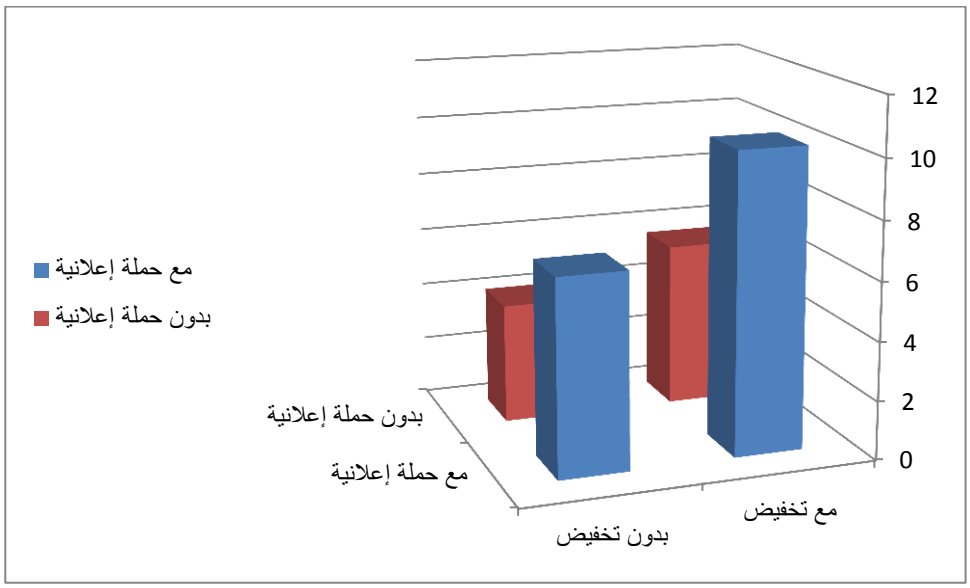
بالتقر على (موافق– OK) نعود مجددا إلى نافذة (تحديد مصدر البيانات – Select Data Source)، ننقر في هذه النافذة على (تحرير – Edit) ضمن الإطار (Horizontal (Category) Axis Labels – مسميات (تصنيفات) المحور الأفقي) لتظهر النافذة الآتية (الشكل 74–ج):

الشكل 74 – ج



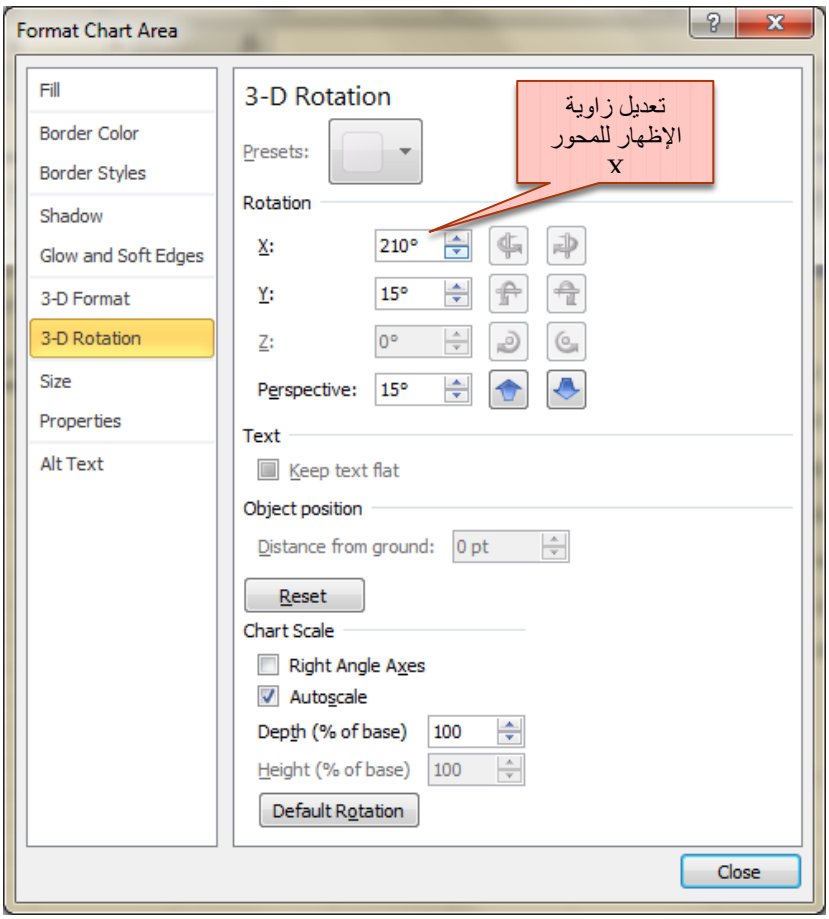
بالنقر على (موافق– OK) لنعود للنافذة (تحديد مصدر البيانات – Select Data Source) والتي ننقر فيها أيضا على (موافق– OK) لنحصل على المخطط الآتي (الشكل 75):

الشكل 75



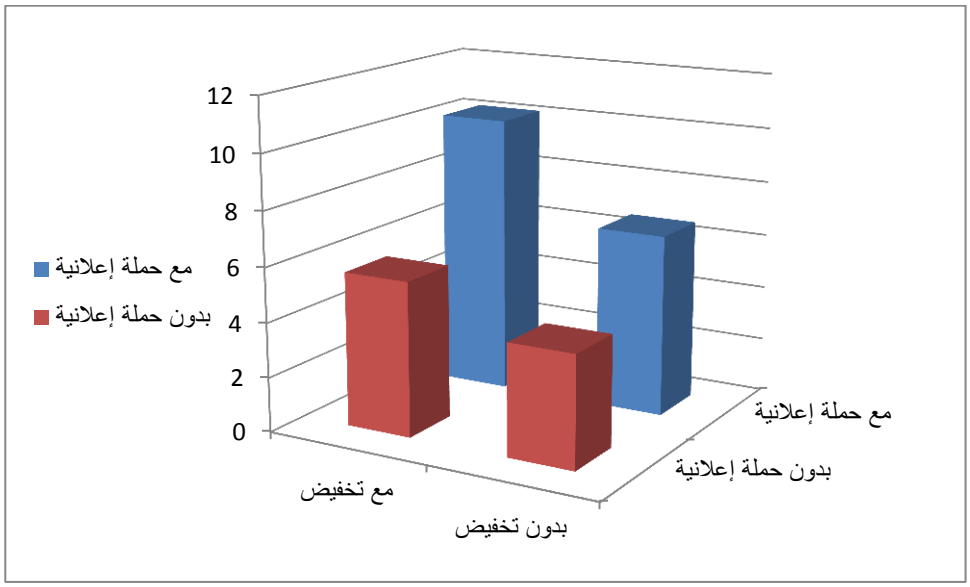
بالطبع، من الأفضل تدوير المخطط حتى تكون الأعمدة القصيرة في مقابل الرؤية، بالنقر على الزر الأيمن للفأرة فوق منطقة ما في المخطط تظهر لدينا قائمة، ننقر فيها على (التدوير ثلاثي الأبعاد – 3D Rotation) لتظهر لدينا النافذة الآتية (الشكل 75):

الشكل 75



بتعديلك على قيم زوايا الروؤية تشهد هذا التغير مباشرة على المخطط، إذا جعلت زاوية الإظهار للمحور X تساوي 120 سيظهر المخطط كما يوضح الشكل 76 الآتي:

الشكل 76

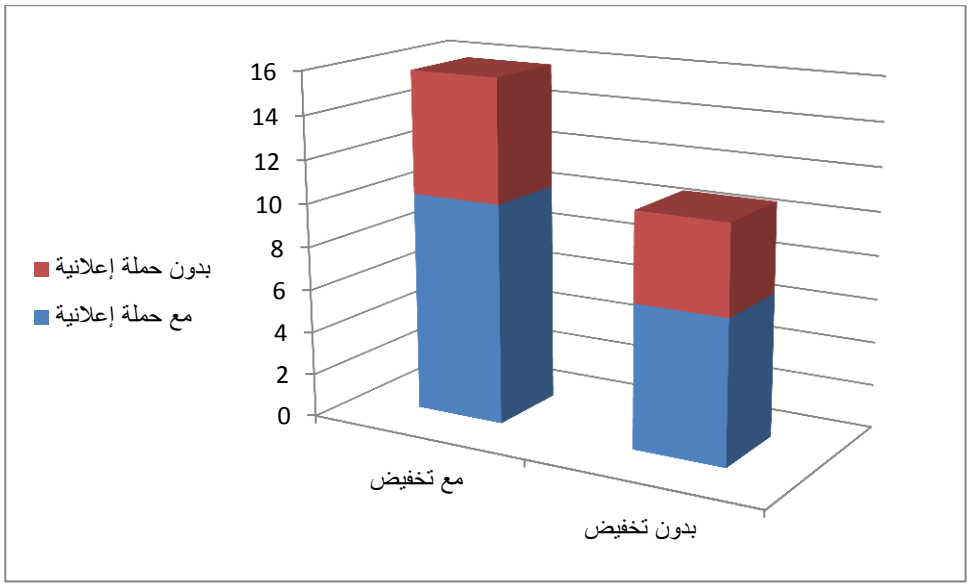


تستطيع تغيير شكل هذا المخطط مباشرة من خلال الخطوات الموضحة في الشكل 77 الآتي:

الشكل 77

ليظهر الشكل 78 الآتي:

الشكل 78

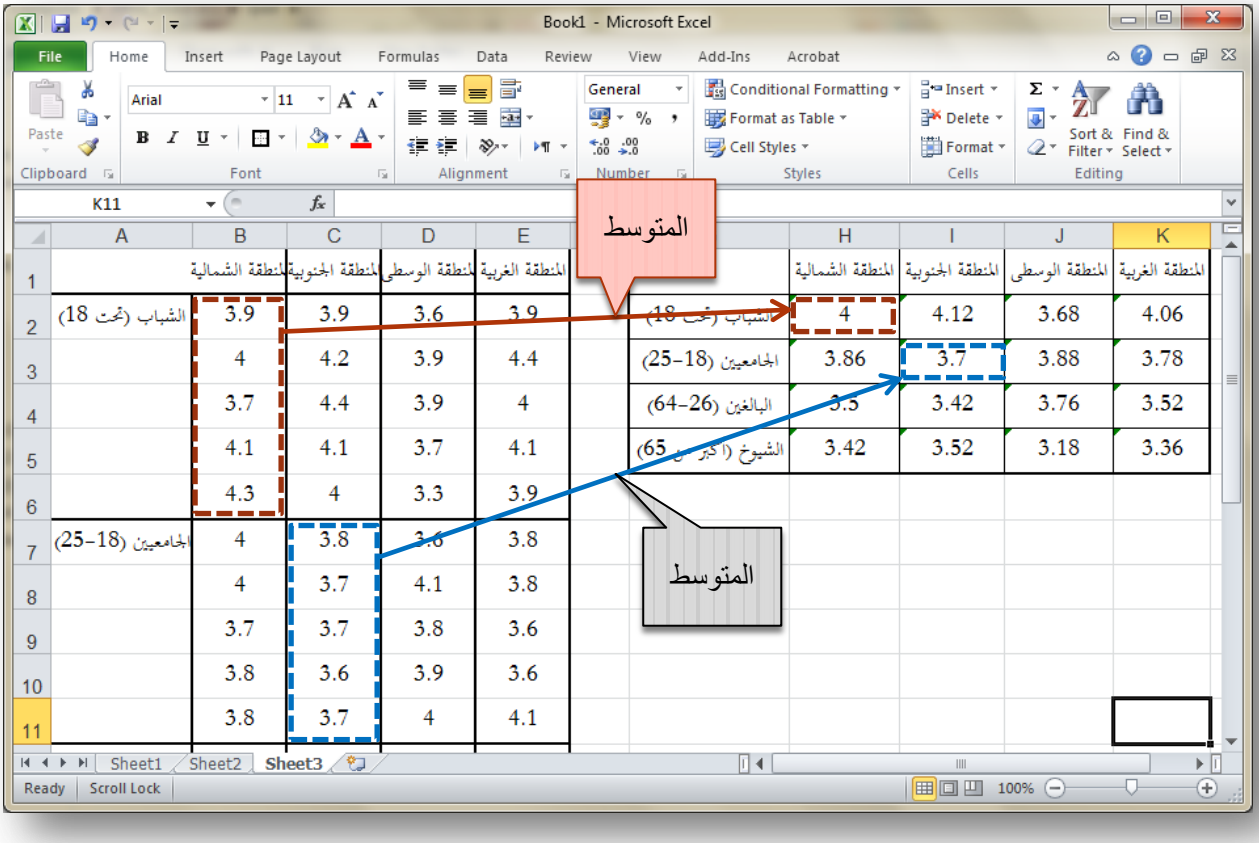


وبسهولة تستطيع اختيار الشكل الجديد الذي ترغب.

2.5 مخططات الخطوط Line Charts:

تتشابه الخطوط مع الأعمدة في طريقة إدخال البيانات، ولتكن بيانات الفقرة 3.3 الموضحة في الشكل 32 ، ولنرتب الجدول الآتي الذي يمثل متوسطات كل تصنيف (كما يوضح الشكل 79 الآتي):

الشكل 79

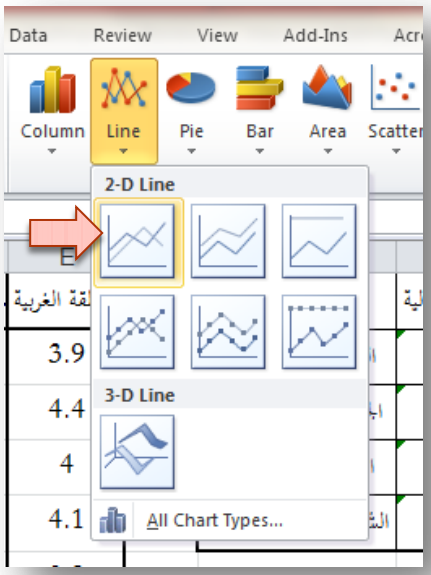


	المنطقة الشمالية	المنطقة الجنوبية	المنطقة الوسطى	المنطقة الغربية
الشباب (تحت 18)	3.9	3.9	3.6	3.9
	4	4.2	3.9	4.4
	3.7	4.4	3.9	4
	4.1	4.1	3.7	4.1
	4.3	4	3.3	3.9
الجامعيين (18-25)	4	3.8	3.6	3.8
	4	3.7	4.1	3.8
	3.7	3.7	3.8	3.6
	3.8	3.6	3.9	3.6
	3.8	3.7	4	4.1

	المنطقة الشمالية	المنطقة الجنوبية	المنطقة الوسطى	المنطقة الغربية
الشباب (تحت 18)	4	4.12	3.68	4.06
الجامعيين (18-25)	3.86	3.7	3.88	3.78
البالغين (26-64)	3.5	3.42	3.76	3.52
الشيوخ (أكبر من 65)	3.42	3.52	3.18	3.36

نضيف المخطط الموضح في الشكل 80 الآتي:

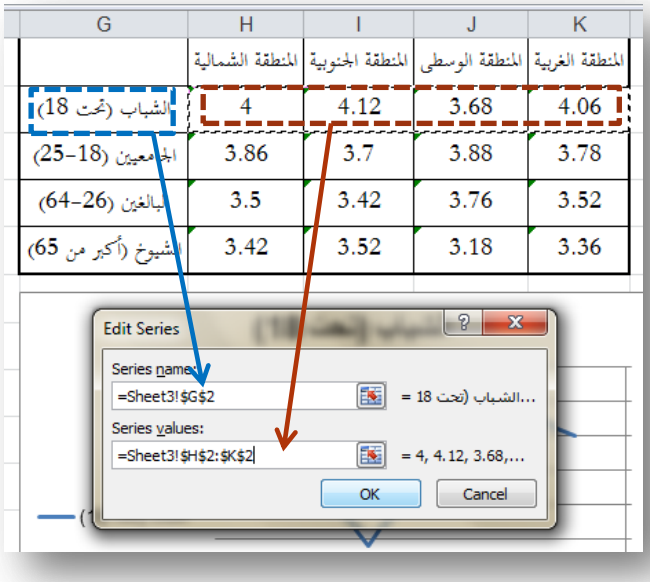
الشكل 80



فيظهر على صفحة إكسل مخطط فارغ، ننقر على هذا المخطط الفارغ بالزر الأيمن للفأرة، نختر من القائمة الظاهرة (تحديد البيانات – Select Data) لتظهر نافذة (تحديد مصدر البيانات – Select Data Source) تماما كما يوضحها الشكل 67 – ب السابق.

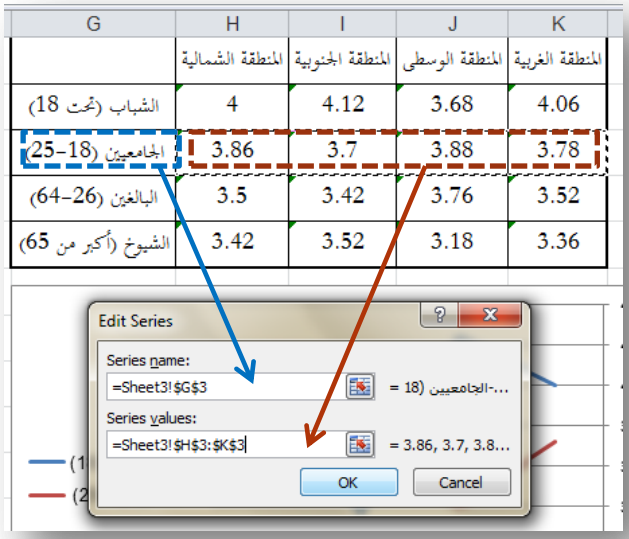
ننقر في هذه النافذة على (إضافة – Add) لتظهر النافذة الآتية (الشكل 80– أ):

الشكل 80 – أ



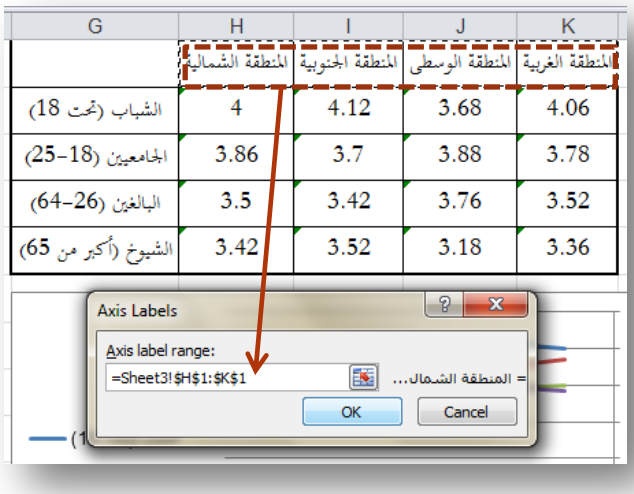
ننقر على (موافق– OK) لنعود للنافذة (تحديد مصدر البيانات – Select Data Source) والتي سنعيد فيها النقر على (إضافة – Add) لتظهر النافذة الآتية (الشكل 80– ب):

الشكل 80– ب



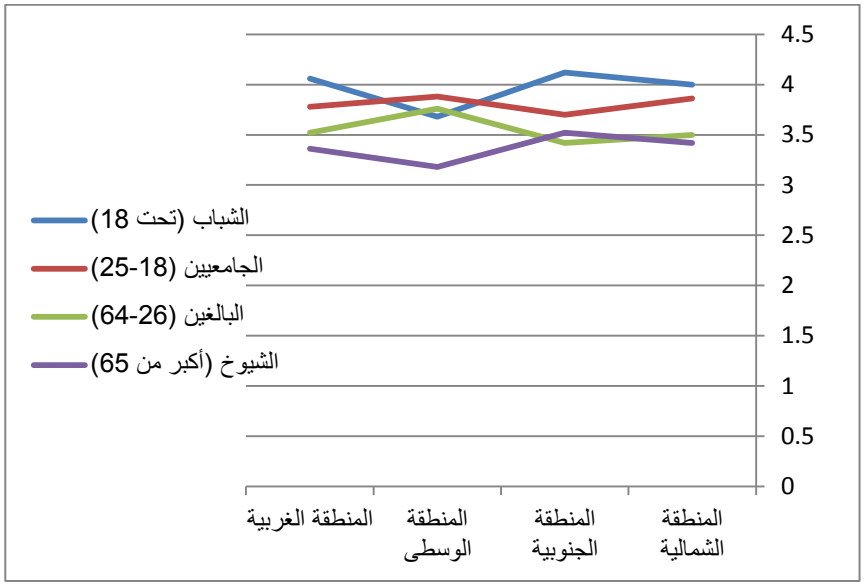
بالتقر على (موافق– OK) نعود مجددا إلى نافذة (تحديد مصدر البيانات – Select Data Source)، نعاود النقر مرة أخرى على (إضافة – Add) لنتابع بقية الأسطر (البالغون والشيوخ). ثم ننقر في هذه النافذة على (تحرير – Edit) ضمن الإطار (Horizontal (Category) Axis Labels – مسميات (تصنيفات) المحور الأفقي) لتظهر النافذة الآتية (الشكل 74–ج):

الشكل 80 – ج



بالنقر على (موافق– OK) لنعود للنافذة (تحديد مصدر البيانات – Select Data Source) والتي ننقر فيها أيضا على (موافق– OK) لنحصل على المخطط الآتي (الشكل 81):

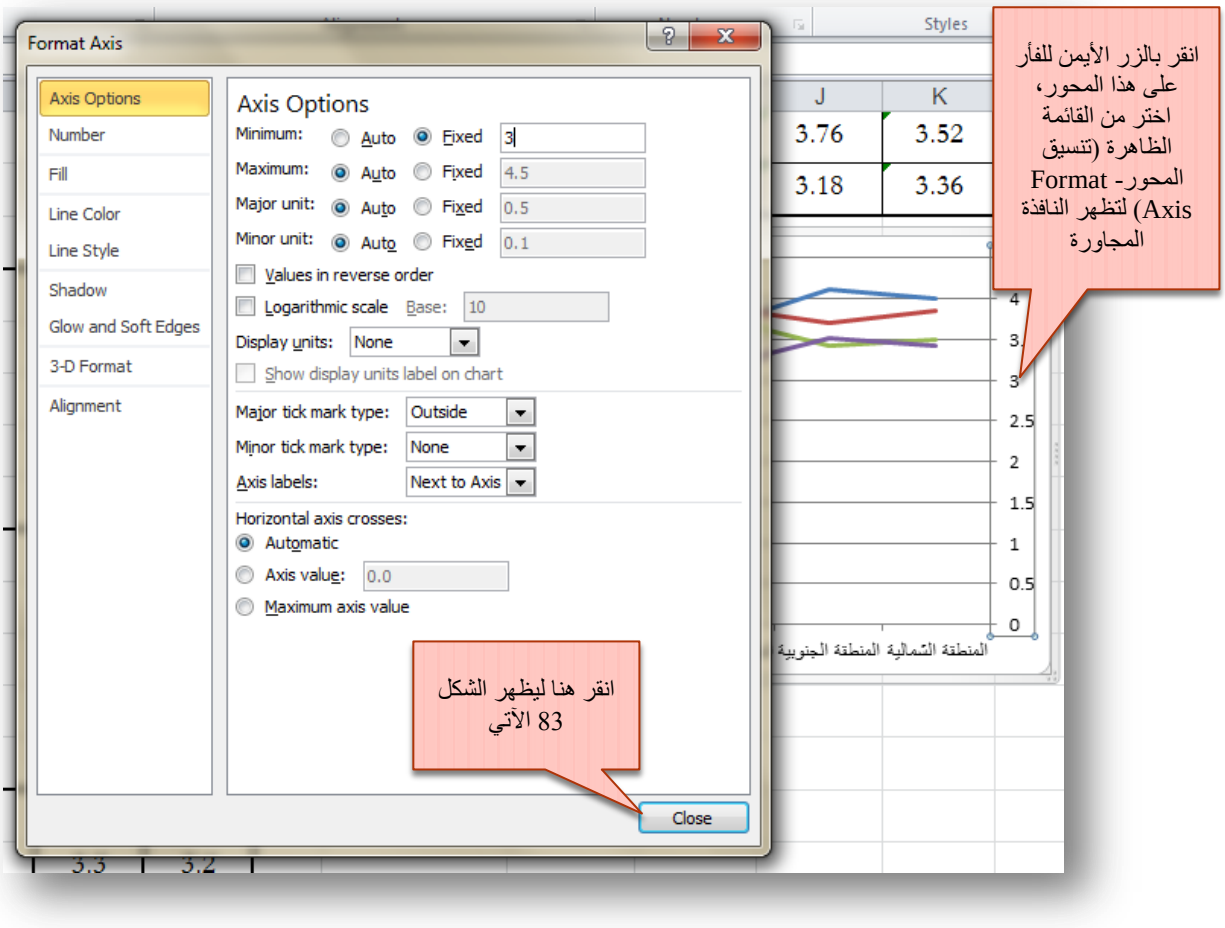
الشكل 81



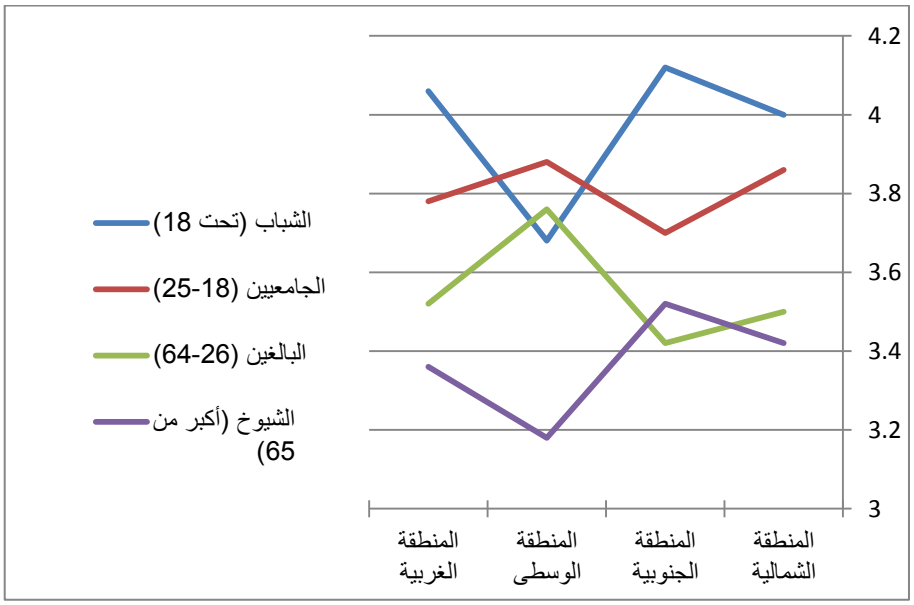
لكن من الواضح أن الخطوط متقاربة من بعضها ولا تعطي صورة واضحة، فما رأيك بتقريب مكان رؤيتك للمخططات بعرض المحور العمودي بدءا من القيمة 3 فقط، وليس بدءا من القيمة 0، أي أننا سنتجاهل المساحة من المخطط التي بين 0 و3 والتي ليس فيها أي رسم، لعمل ذلك اتبع الخطوات الموضحة في الشكل 82 الآتي:

الشكل 82

حدد الخيار (مثبت – Fixed) ، وفي الحقل المجاور القيمة (3)


Format Axis
Axis Options
Number
Fill
Line Color
Line Style
Shadow
Glow and Soft Edges
3-D Format
Alignment
Axis Options
Minimum: Auto Fixed 3
Maximum: Auto Fixed 4.5
Major unit: Auto Fixed 0.5
Minor unit: Auto Fixed 0.1
Values in reverse order
Logarithmic scale Base: 10
Display units: None
Show display units label on chart
Major tick mark type: Outside
Minor tick mark type: None
Axis labels: Next to Axis
Horizontal axis crosses:
Automatic
Axis value: 0.0
Maximum axis value
Close
انقر بالزر الأيمن للفأر على هذا المحور، اختر من القائمة الظاهرة (تنسيق المحور- Format Axis) لتظهر النافذة المجاورة
انقر هنا ليظهر الشكل 83 الآتي
Styles
J
K
3.76
3.52
3.18
3.36
4
3.5
3
2.5
2
1.5
1
0.5
0
المنطقة الشمالية
المنطقة الجنوبية
3.3
3.2

الشكل 83


الشباب (تحت 18)
الجامعيين (18-25)
البالغين (26-64)
الشيوخ (أكبر من 65)
4.2
4
3.8
3.6
3.4
3.2
3
المنطقة الغربية
المنطقة الوسطى
المنطقة الجنوبية
المنطقة الشمالية