الفصل الثاني
الانحدار الخطي البسيط

فروض الخطأ العشوائي:

لقد سبق وذكرنا أن الدور الأساسي للاقتصاد القياسي ينحصر في تقدير معالم النموذج الاقتصادي من خلال مشكلة عدم التأكد (Uncertainty - Incertitude). وذكرنا أن النموذج الاقتصادي قد يتضمن العديد من المعادلات، كما وقد تتضمن كل معادلة على الكثير من المتغيرات. ويتوقف عدد المعادلات وعدد المتغيرات التي يتضمنها أي نموذج اقتصادي على طبيعة الظاهرة الاقتصادية التي يرمي الباحث إلى تفسير اختلافاتها. والجدير بالذكر، أن معالجة مثل هذه النماذج الاقتصادية المعقدة يصبح أمراً بسيطاً، فيما لو أتقن الباحث معالجة النماذج الاقتصادية البسيطة التي تتناول تفسير العلاقة بين متغيرين أحدهما تابع (Dependent variable – variable dépendante) والآخر مستقل (Independent variable – Variable indépendante). ومثال ذلك تفسير العلاقة بين الاستهلاك الشخصي (المتغير التابع C_t)، والدخل الشخصي المتاح (المتغير المستقل Y_t) والتي يمكن صياغتها في نموذج بسيط كالآتي:

$$C_t = \alpha_0 + \alpha_1 Y_t \pm U_t$$

تشير الصيغة أعلاه إلى حقيقة هامة، وهي أن العلاقات بين المتغيرات الاقتصادية، غير تامة (Inexact relationship – Relation exacte)، حيث يقيس المتغير العشوائي U_t أثر متغيرات أخرى ـ غير الدخل المتاح ـ لم نتمكن من قياسها وإدخالها بشكل صريح في المعادلة. فلو فرضنا مثلاً أنه توافر لدينا بيانات دقيقة وكاملة عن كل من الاستهلاك الشخصي والدخل الشخصي المتاح لمجتمع إحصائي مكوّن من 1000 عائلة. وأننا استطعنا تصنيف هذا المجتمع في مجتمعات فرعية (Subpopulations – Sous populations) بحيث يتساوى الدخل الشخصي المتاح للأسر ضمن كل مجتمع فرعي. فإننا نلاحظ في هذه الحالة، أن المتغير المستقل (الدخل المتاح) أصبح ثابتاً (Fixed variable - variable fixe) ضمن كل مجتمع فرعي. فإذا فرضنا أن المجتمع الفرعي الأول يتضمن 170 عائلة تتساوى في دخلها المتاح وليكن 10000$ في السنة. بينما يتضمن المجتمع الفرعي الثاني كل العائلات التي تحصل على نفس مستوى الدخل المتاح وليكن 12000$ في السنة، وهكذا، فإننا نحصل على مجتمعات فرعية، تتميز عن بعضها بتساوي الدخل المتاح لأفراد المجتمع الفرعي الواحد.

دعنا الآن نفترض أنه في دراساتنا للعلاقة بين الاستهلاك والدخل، سنكتفي بالبيانات المتوفرة عن مجتمع فرعي معين. فلو أخذنا المجتمع الفرعي الأول فإننا سنلاحظ، أنه بالرغم من تساوي مستوى الدخل الشخصي المتاح لكل هذه العائلات، إلا أنها سوف تظهر تفاوتاً في إنفاقها الاستهلاكي. ويعود السبب في ذلك إلى عوامل أخرى غير الدخل المتاح كاختلاف حجم وتركيب كل عائلة، واختلاف هوايات وطباع رب المنزل من حيث كونه مثقف، رياضي، اجتماعي، أو مضياف . . . . وبذلك نخلص إلى أنه لا يوجد علاقة تامة بين الدخل والاستهلاك، حيث يمكن لعائلتين (أو أكثر) أن تختلفا في إنفاقهما الاستهلاكي على الرغم من تساوي دخلهما الشخصي المتاح.

الجدير بالذكر، أنه بالرغم من ثبات الدخل المتاح ضمن كل مجتمع فرعي، إلا أنه توجد علاقة حقيقية في المعدل (On the average - n moyennee) بين الاستهلاك الشخصي والدخل الشخصي المتاح. فلو رمزنا إلى القيمة المتوقعة في المعدل (Expected value – Valeur espérée) للإنفاق الاستهلاكي عند الدخل الشخصي المتاح لكل العائلات في المجتمع الفرعي،

$$E(C/Y) = \mu_{cy}$$

$$E(C/Y) = \mu_{cy} = \alpha + \beta Y$$

حيث تمثل μ_{cy} العلاقة الحقيقية المتوقعة في المعدل (الوسط الحسابي) بين الاستهلاك الشخصي والدخل الشخصي المتاح لكل العائلات ضمن المجتمع الفرعي. أما الإنفاقات الاستهلاكية الفردية للعائلات ضمن المجتمع الفرعي الواحد، فستكون مجمعة (Clustered - Régroupées) حول هذه القيمة. وبمعنى أدق، فباستطاعتنا مثلاً أن نمثل العلاقة بين الدخل الشخصي المتاح والاستهلاك الشخصي للعائلة الأولى في المجتمع الفرعي الأول، بالدالة التالية:

$$C = \alpha + \beta Y + U_1$$

وأن نمثل هذه العلاقة، للعائلة الثانية في المجتمع الفرعي الأول بالدالة التالية:

$$C = \alpha + \beta Y + U_2$$

وأن نمثل هذه العلاقة، للعائلة الأخيرة في المجتمع الفرعي الأول بالدالة التالية:

$$C = \alpha + \beta Y + U_k$$

بحيث ترمز U_1، U_2...... U_k إلى الخطأ العشوائي الذي يقيس اختلاف العلاقة بين الدخل والاستهلاك عن العلاقة الحقيقية للمجتمع الفرعي الواحد. فيختلف استهلاك العائلة الأولى عن القيم المتوقعة للاستهلاك في المجتمع الفرعي الأول بمقدار U_1، بينما يختلف استهلاك العائلة الثانية في المجتمع الفرعي الأول عن استهلاك المتوقع في هذا المجتمع بمقدار U_2 الخ. آخذين في الاعتبار أنه توجد عوامل أخرى غير الدخل المتاح، تؤثر في الإنفاق الاستهلاكي لكل عائلة ضمن المجتمع الفرعي، فتجعل استهلاك أية عائلة ضمن هذا المجتمع الفرعي، مساوياً أو أكبر أو أقل، من القيمة المتوقعة في المعدل للإنفاق الاستهلاكي لكل العائلات في المجتمع الفرعي الواحد. ونظراً لأن هذه العوامل الأخرى قد تسير في اتجاهات مختلفة، فتأثر بشكل مختلف على الإنفاقات الفردية للعائلات، لذلك فإننا نفترض عادة، أن المتغير العشوائي U يتوقف على عامل الصدفة بحيث أن U_1 قد تكون أقل أو أكبر أو مساوياً للصفر، وكذلك الحال بالنسبة لقيمة U_2 و U_3 و .....U_k. لكن الوسط الحسابي (Arithmetic Mean - Moyenne arithmétique) لهذه القيم سوف يساوي صفراً. ونظراً لأن المتغير العشوائي U هو متغير غير منتظم، لذلك لا بدّ من صياغة فروض خاصة به (Assumptions underlying the error term – Hypothèses concernant le terme d'erreur) تمكنا من إدخاله في النموذج الاقتصادي، وحتى نتمكن من إيضاح هذه الفروض بشكل مبسط سنأخذ المثال الآتي:

تقترح النظرية الاقتصادية وجود علاقة دالية بين الكمية المعروضة من سلعة ما وسعر هذه السلعة في السوق. دعنا نفترض أن المتغير المستقل (السعر) هو متغير ثابت في المجتمعات الفرعية، في حين نفترض أن المتغير التابع (العرض) هو متغير عشوائي (Random variable – Variable aléatoire) في المجتمعات الفرعية.

بمعنى أننا سنفترض أنه لو ذهبنا إلى السوق واخترنا سعراً ثابتاً وليكن عشرة قروش مثلاً، فإننا سنلاحظ أن الكمية المعروضة من السلعة عند هذا السعر الثابت وفي إحدى المتاجر ستكون متفاوتة

دعنا نفترض أن السوق الكلي يتضمن خمسة مؤسسات تعرض السلعة Z، ودعنا نفترض أيضاً، أننا تمكنا من تسجيل الكمية المعروضة من السلعة Z، في هذه المؤسسات عندما كان سعر السلعة في السوق يساوي (10) قروش، ثم سجلنا الكمية المعروضة عند السعر الثابت في السوق وقدره (11) قرش، ثم عند السعر الثابت (12) قرش و. . . . الخ، كما هو مبين في الجدول رقم (1):

الجدول رقم (1)
سعر السلعة في خمسة أسواق مختلفة،
والكمية المعروضة من السلعة Z في خمسة مؤسسات مختلفة ضمن
السوق الواحد

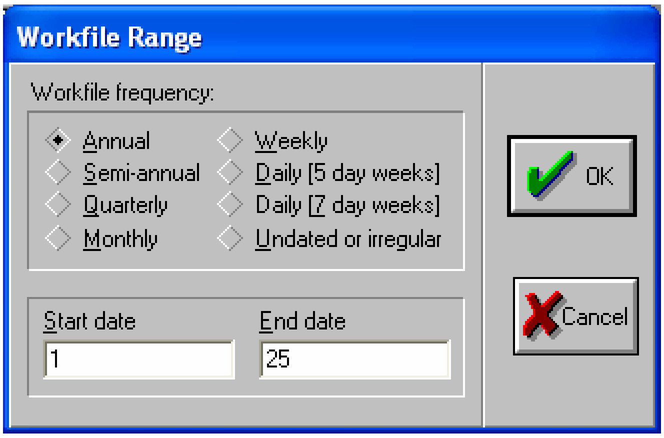
المجتمع الفرعي	السعر الثابت X	الكمية المعروضة في السوق الواحد عند السعر الثابت وفي مؤسسات مختلفة Y	الوسط الحسابي للكمية المعروضة عند السعر الثابت، في السوق الواحد E(Y/X)
الأول	10	34 30 32 31 33	32
الثاني	11	36 33 34 36 31	34
الثالث	12	42 38 31 34 35	36
الرابع	13	45 33 37 36 39	38
الخامس	14	40 39 51 36 34	40

نلاحظ في الجدول الفرضي أعلاه أن تصنيف البيانات قائم على أساس أن السعر الثابت في السوق. ففي السوق الأول مثلاً، كان السعر للسلعة Z يساوي (10) قرشاً، وكانت الكمية المعروضة من السلعة Z في نفس الوقت، وفي خمسة مؤسسات مختلفة، كالآتي:

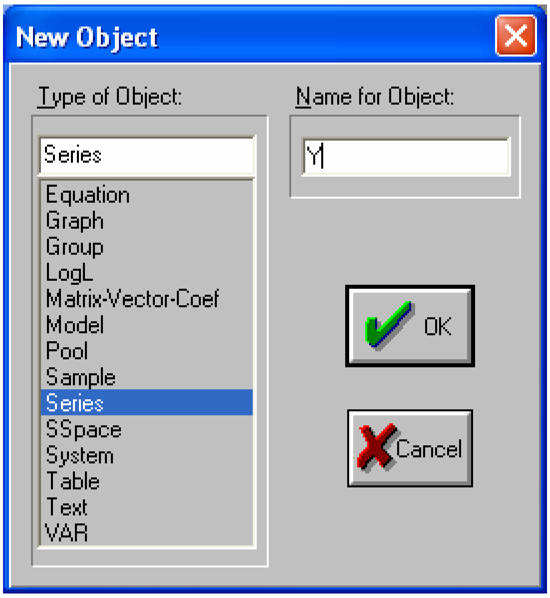
34 ، 30 ، 32 ، 31 ، 33

أما الوسط الحسابي للكمية المعروضة من السلعة Z في السوق فهو 32. فعلى الرغم من أن سعر السلعة Z في السوق كان (10) قرشاً إلا أن الكميات المعروضة في السوق كانت متفاوتة حول الوسط الحسابي 32، كذلك كانت الكميات المعروضة من السلعة في السوق الثاني عند السعر الثابت (11) قرشاً، مجمعة حول الوسط الحسابي 34 وهكذا. ويمكننا إيضاح العلاقة بين القيم الفعلية المشاهدة عن الكمية المعروضة والسعر لبيانات الجدول رقم (1) بيانياً كالآتي:

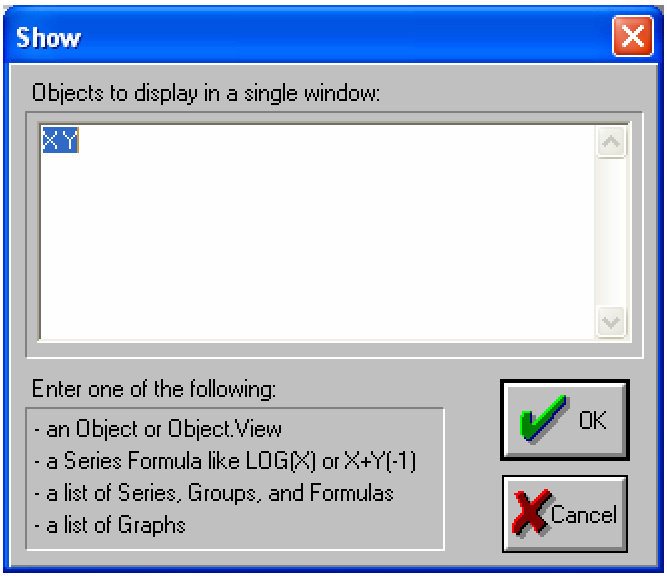
FILE \ NEW \ WORKFILE

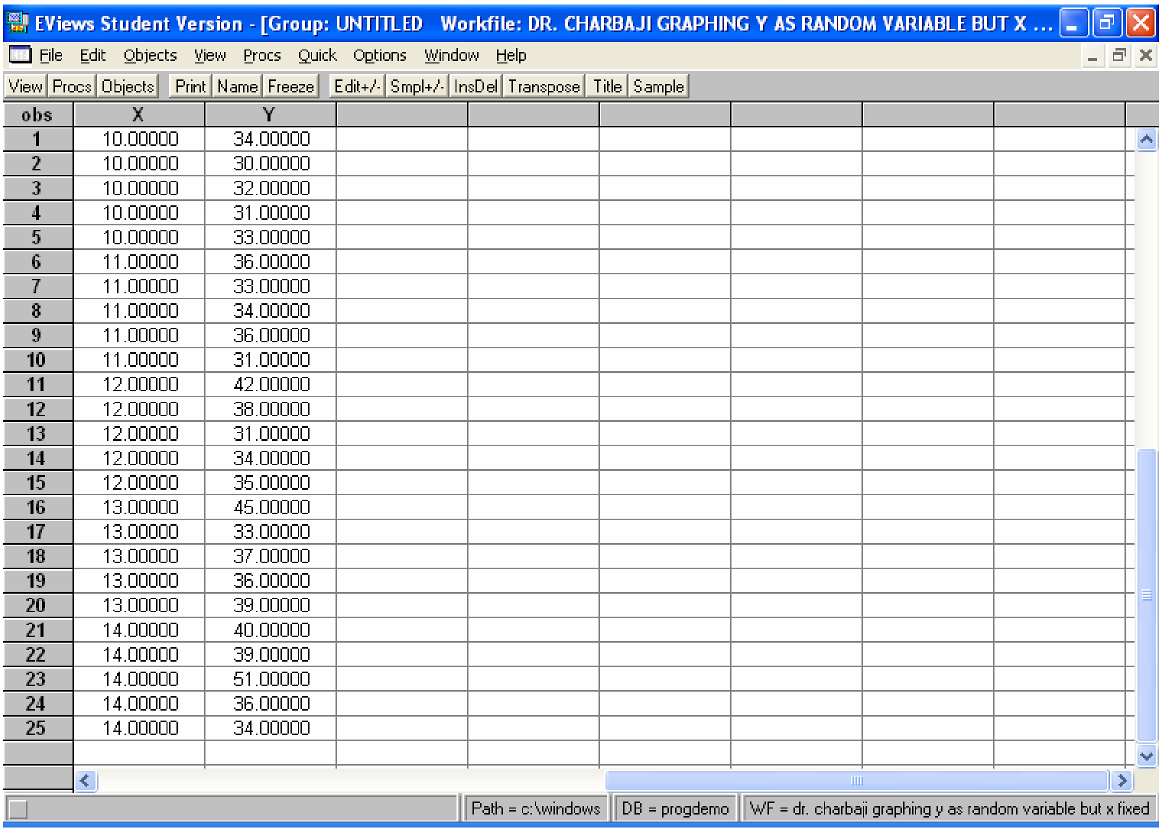


OBJECT / NEW OBJECT / SERIES

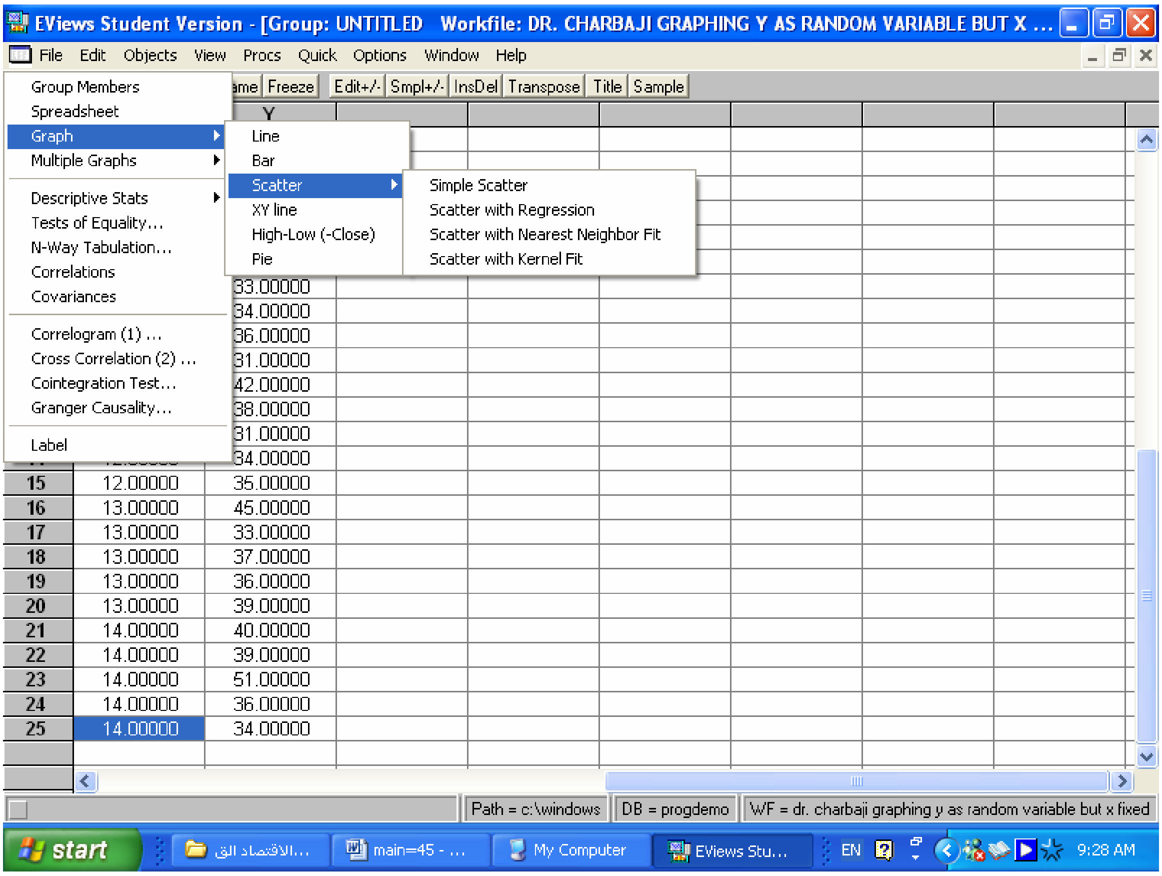


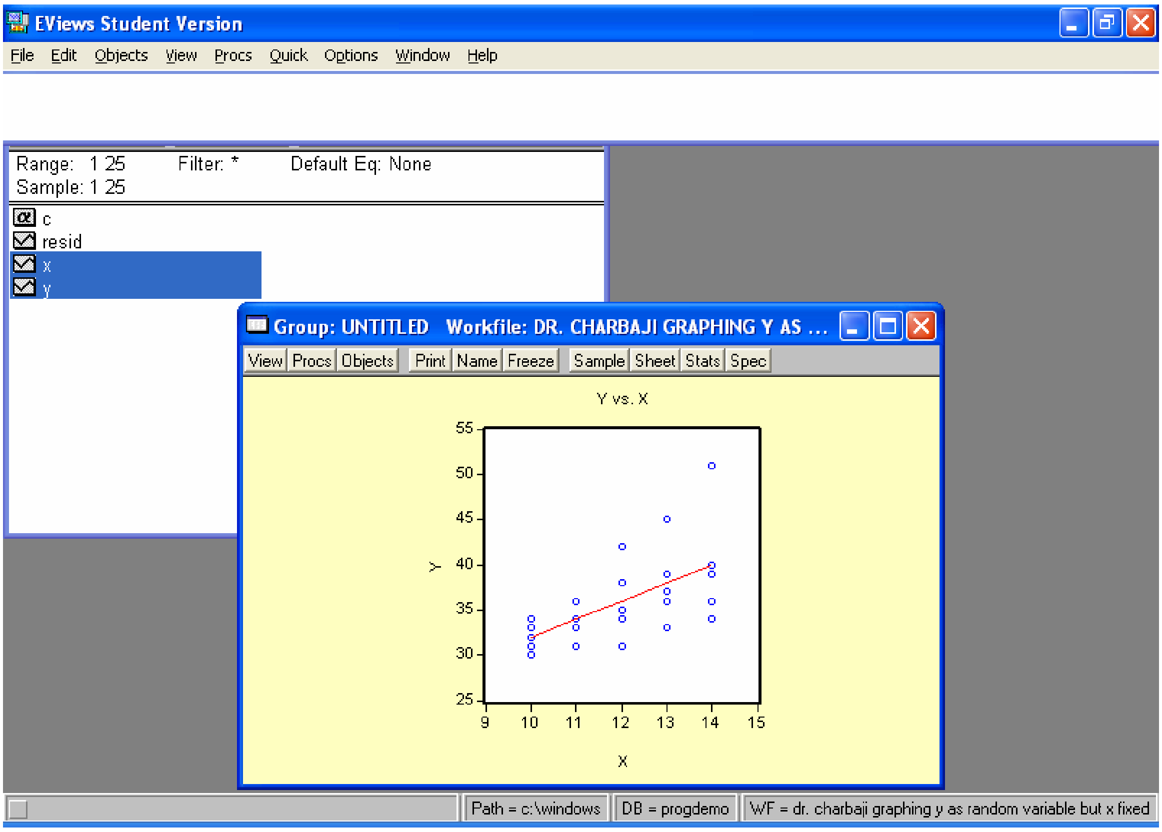
VIEW / SHOW



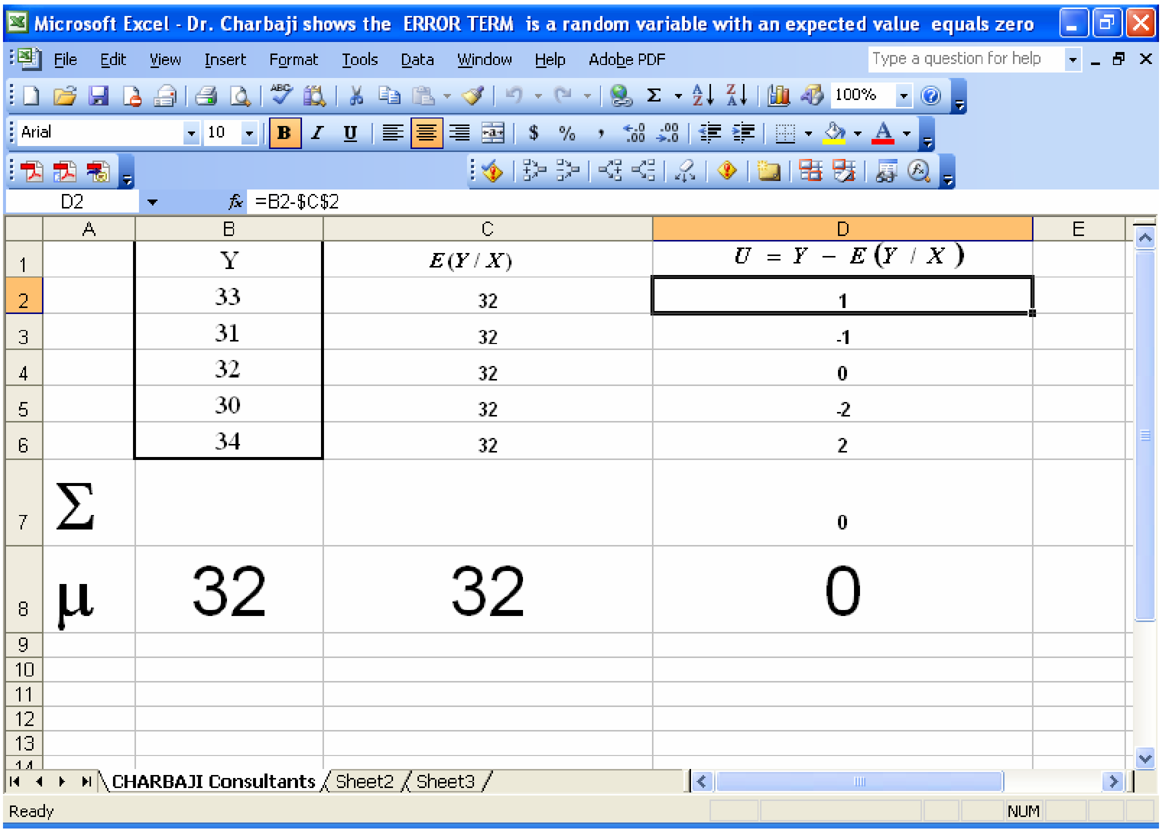


obs	X	Y
1	10.00000	34.00000
2	10.00000	30.00000
3	10.00000	32.00000
4	10.00000	31.00000
5	10.00000	33.00000
6	11.00000	36.00000
7	11.00000	33.00000
8	11.00000	34.00000
9	11.00000	36.00000
10	11.00000	31.00000
11	12.00000	42.00000
12	12.00000	38.00000
13	12.00000	31.00000
14	12.00000	34.00000
15	12.00000	35.00000
16	13.00000	45.00000
17	13.00000	33.00000
18	13.00000	37.00000
19	13.00000	36.00000
20	13.00000	39.00000
21	14.00000	40.00000
22	14.00000	39.00000
23	14.00000	51.00000
24	14.00000	36.00000
25	14.00000	34.00000


EViews Student Version - [Group: UNTITLED Workfile: DR. CHARBAJI GRAPHING Y AS RANDOM VARIABLE BUT X ...
File Edit Objects View Procs Quick Options Window Help
Group Members
Spreadsheet
Graph
Multiple Graphs
Descriptive Stats
Tests of Equality...
N-Way Tabulation...
Correlations
Covariances
Correlogram (1) ...
Cross Correlation (2) ...
Cointegration Test...
Granger Causality...
Label
Line
Bar
Scatter
XY line
High-Low (-Close)
Pie
Simple Scatter
Scatter with Regression
Scatter with Nearest Neighbor Fit
Scatter with Kernel Fit
Path = c:\windows
DB = progdemo
WF = dr. charbaji graphing y as random variable but x fixed


EViews Student Version
File Edit Objects View Procs Quick Options Window Help
Range: 1 25 Filter: * Default Eq: None
Sample: 1 25
c
resid
x
y
Group: UNTITLED Workfile: DR. CHARBAJI GRAPHING Y AS ...
Y vs. X
Y
X
Path = c:\windows
DB = progdemo
WF = dr. charbaji graphing y as random variable but x fixed

حيث ترمز Y الى القيمة الفعلية المشاهدة، في حين ترمز $Y^{\backslash}$ إلى القيمة المتوقعة $E(Y/X)$ ويطلق على انحراف (Deviation -Deviation) القيمة الفعلية المشاهدة Y (The observed value – La valeur observée) عن القيمة المتوقعة في المعدل للمجتمع الفرعي $E(Y/X)$ اسم الخطأ العشوائي، ويمكن كتابة القيم الفعلية المشاهدة على أنها مكونة من جزأين: الجزء الأول وهو عبارة عن القيمة المتوقعة، والجزء الثاني وهو عبارة عن الخطأ العشوائي، وبالتالي يمكننا كتابة القيم الفعلية المشاهدة للكمية المعروضة في السوق الأول كالآتي:



	A	B	C	D
1		Y	$E(Y/X)$	$U = Y - E(Y/X)$
2		33	32	1
3		31	32	-1
4		32	32	0
5		30	32	-2
6		34	32	2
7	Σ			0
8	μ	32	32	0

الجدير بالذكر، أننا افترضنا وللتبسيط في الشرح، إن السوق الكلي يتكون من خمسة مجتمعات فرعية في كل منها خمسة مشاهدات. إلا إن واقع الحياة الاقتصادية مختلف تماماً، حيث يتكون المجتمع الإحصائي من عدد لا نهائي من المجتمعات الفرعية ومن المشاهدات (observation - observation) . كما وإن القيمة المتوقعة $\mu_{xy} = E(Y/X)$، هي في الحقيقة قيمة نظرية لا يمكن للباحث قياسها.

فالباحث عادة يحصل على أزواج من المشاهدات (Pairs of observations - Paires d'observations) $X_1Y_1, X_2Y_2, X_3Y_3, ... X_KY_K$ في عيّنة من البيانات ، ويعتبر إن كل زوج من هذه المشاهدات هو عيّنة عشوائية ممثلة تمثيلاً صادقاً للمجتمع الفرعي (Representative random sample – échantillon aléatoire représentatif)، ومن ثم فإنه يعمل على تقدير العلاقة الحقيقية في المجتمع الإحصائي من خلال بيانات العينة، مفترضاً بذلك فروضاً خاصة بالمتغير العشوائي، وهي:

أولاً: نفترض أن الخطأ العشوائي هو متغير عشوائي مستقل (Independent random variable – Variable aléatoire indépendante) وتعتمد قيمه على عامل الصدفة، فقد يكون

$$Y = \alpha + \beta X + U$$

$$U = Y - \alpha - \beta X$$

$$\sum U = \sum (Y - \alpha - \beta X)$$

$$\sum U = \sum Y - \sum \alpha - \beta \sum X$$

$$\sum U = \sum Y - n\alpha - \beta \sum X$$

$$\sum U = \sum Y - n(\overline{Y} - \beta \overline{X}) - \beta \sum X \ ^{*}$$

$$\sum U = \sum Y - n\overline{Y} + \beta n \overline{X} - \beta \sum X$$

$$\sum U = \sum Y - \sum Y + \beta \sum X - \beta \sum X$$

$$\sum U = 0$$

$$E(U) = 0$$

ثانياً: نفترض أن للمتغير العشوائي نفس التباين (Variance - Variance) في المجتمعات الفرعية، كما وان هذا التباين الثابت (Constant variance – Variance constante) للمتغير العشوائي يساوي تباين المتغير التابع y في المجتمع الإحصائي. ويمكن إيضاح هذه الفرضية كالآتي :

$$Y = E(Y / X) + U$$

$$\sigma_y^2 = E[Y - \mu_{yx}]^2$$

$$\sigma_y^2 = E[Y - E(Y / X)]^2$$

$$\sigma_y^2 = E(U)^2$$

$$\sigma_y^2 = \sigma_u^2 = \sigma_{yx}^2$$

---

[1] Aigner, p:25

* لاحظ أنّ $\alpha = \overline{Y} - \beta \overline{X}$. أنظر في ذلك طريقة المربعات الصغرى.

ثالثاً: نفترض أن المتغير العشوائي يتوزع في كل مجتمع فرعي بشكل التوزيع المعتدل (Normally distributed – distribué normalement) حول القيمة المتوقعة E(Y/X). علماً بان الهدف من صياغة هذه الفرضية هي إمكانية استخدام الاختبارات الإحصائية (F and t-tests فيما بعد). ذلك انه لا يمكن استخدام هذه الاختبارات إلا على بيانات لمجتمع إحصائي موزع توزيعاً معتدلاً*.

ويمكن صياغة الفروض الثلاثة السالفة الذكر كالآتي:

$$U \sim N(0, \sigma^2)$$

بمعنى أن U تتوزع بشكل معتدل، حيث أن الوسط الحسابي للتوزيع يساوي صفراً كما وان التباين ثابت .

رابعاً: نفترض انعدام التغيير (covariance – covariance) بين المتغير العشوائي U والمتغير المستقل X ويمكن إيضاح هذه الفرضية كالآتي[2]:

$$\sum XU = \sum X\,(Y - \alpha - \beta X)$$

$$\sum XU = \sum XY - \alpha \sum X - \beta \sum X^2$$

$$\sum XU = \sum XY - \sum X\,(\bar{Y} - \beta\,\bar{X}) - \beta \sum X^2$$

$$\sum XU = \sum XY - \bar{Y}\sum X + \beta\bar{X}\sum X - \beta \sum X^2$$

$$\sum XU = \left[\sum XY - \frac{(\sum Y)(\sum X)}{n}\right] - \beta\left[\sum X^2 - \frac{(\sum X)(\sum X)}{n}\right]^{*}$$

$$\sum XU = \sum xy - \frac{\sum xy}{\sum x^2}.\sum x^2$$

---

* لاحظ أن $\grave{Y} = b_0 + b_1\,x + e$، بمعنى أن Y هي التركيب الخطي للمتغير العشوائي، وبما أن e تتوزع توزيعاً معتدلاً، لذلك فإن Y، وبالتالي b_0 و b_1 تتوزع بشكل معتدل أيضاً، علماً أن e هي تقدير لقيمة U من بيانات العينة.

[2] التغاير هو القيمة المتوقعة في المعدل لحوا صل ضرب انحرافات قيم المتغيرين عن الوسط الحسابي $Cov(X,Y) = \frac{\sum xy}{n}$ ويستخدم في قياس الارتباط بين المتغيرين. انظر في ذلك معامل الارتباط .

أما البرهان الرياضي أعلاه فهو مقتبس من Aigner ص:25. علماً أن : $x = X - \bar{X}\ ,\ \ y = Y - \bar{Y}$

* لاحظ أنّ:

$$\sum (X - \bar{X})(Y - \bar{Y}) = \sum xy = \sum XY - \frac{\sum X \sum Y}{n}$$

$$\sum (X - \bar{X})^2 = \sum x^2 = \sum X^2 - \frac{(\sum X)^2}{n}$$

$$\beta = \frac{\sum xy}{\sum x^2}$$

كذلك تجدر الإشارة إلى أنّ التغاير بين الخطأ العشوائي والقيم المتوقعة يساوي صفراً, وذلك بالاعتماد على الفرضية أولاً ورابعاً أعلاه:

$$\sum \grave{Y} e = \sum (b_0 + b_1 X)e = b_0 \sum e + b_1 \sum X e = 0$$

$\sum XU = 0$

خامساً: نفترض انعدام العلاقة بين قيم المتغير العشوائي U_t في الفترة t، وقيمه في الفترة السابقة U_{t-1} أو اللاحقة U_{t+1}.

سادساً: نفترض عدم وجود أخطاء في قياس المتغير المستقل.

سابعاً: نفترض انعدام العلاقة القوية (Multicollinearity - Multicolinéarité) بين المتغيرات المستقلة. وهي فرضية تفيد عند استخدام الانحدار المتعدد.

ثامناً: نفترض أنه لا يوجد أخطاء في تحديد النموذج. (There is no specification – error - Il n'y a pas d'erreur de spécification).

تاسعاً: نفترض أنه تم تجميع البيانات الاقتصادية في شكلها الكلي (Aggregation - Agrégation) بشكل صحيح.

إذن نخلص إلى أننا لا نعمل في الاقتصاد القياسي على توقع (تنبؤ) قيمة معينة للمتغير التابع وإنما نعمل على تقدير (Estimation - estimation) القيمة المتوقعة في المعدل $E(Y/X)$ للمجتمع الإحصائي، ونعتمد في ذلك على البيانات الاقتصادية الفعلية والتي تعتبر عينة عشوائية ممثلة للمجتمع الإحصائي بأكمله. ويمكن تلخيص ما سبق ذكره كالآتي:

العلاقة الحقيقية (The true relationship – La vraie relation) بين المتغيرين X and Y في المجتمع الإحصائي :

$$Y = \alpha + \beta X + U$$

العلاقة المقدرة (The estimated relationship – la relation estimée) بين المتغيرين X and Y من بيانات العينة :

$$\hat{Y} = b_0 + b_1 X + e$$

معادلة الانحدار الحقيقي (The true regression equation – La vraie équation de régression) بين المتغيرين X and Y في المجتمع الإحصائي :

$$E(Y/X) = \alpha + \beta X$$

معادلة الانحدار المقدّرة (the estimated regression equation – L'équation de regression estimée) بين المتغيرين X and Y من بيانات العينة:

$$\hat{Y} = b_0 + b_1 X$$

أما المشكلة الأساسية في الاقتصاد القياسي فتنحصر في إيجاد قيم معاملات الانحدار (The regression coefficient) b_0 and b_1 والتي تعتبر تقديرات (estimates - estimations) لقيم المعالم (parameters - paramètres) β , α في المجتمع الإحصائي.

(The method of least squares – طريقة المربعات الصغرى Méthode des moindres carrés ordinaires)

تنص نظرية جاوس- ماركوف (The Gauss – Markov theorem – Théorème de Gauss-Markov) على أنه إذا كان المتغير المستقل ثابتاً والمتغير التابع عشوائياً في المجتمعات الفرعية ، المتساوية التباين ، فإن طريقة المربعات الصغرى تعطي أكفأ تقديرات خطية غير متحيزة (BLUE=Best linear unbiased estimators – Meilleurs estimateurs sans biais) علماً أن طريقة المربعات الصغرى ،هي الطريقة التي ترشدنا إلى إيجاد قيم b_1 و b_0 والتي تجعل مجموع مربعات البواقي عند نهايتها الصغرى.

دعنا نفترض أنه لدينا n من أزواج المشاهدة (Pairs of observations - Paires d'observations) $X_1Y_1, X_2Y_2, ...X_nY_n$ وأننا نريد أن نوفق أفضل خط للعلاقات بين الظاهرتين X and Y (The best fitting line – La droite la plus adéquate).

نلاحظ انه بالنسبة لأية قيمة في X ولتكن X_1 مثلاً ، فإن القيمة الفعلية المشاهدة Y_1 ستختلف عن القيمة المتوقعة $Y_1^{\backslash}$ والواقعة على خط الانحدار (The regression line – La droite de regression) بالمقدار e_1 والذي يعرف بالبواقي (Residual - Résidu). ولو أننا أخذنا كل قيم X وأوجدنا الاختلافات $e_1, e_2 .....e_n$ فسنجد أن بعض هذه الاختلافات موجب ،وبعضها سالب وبعضها الأخر يساوي صفراً. أما مجموع قيم الاختلافات $\sum(Y - Y^{\backslash}) = \sum e = 0$ فيساوي صفراً. وبإمكاننا أن نقيس مدى جودة توفيق خط العلاقات وذلك بأخذ مجموع مربع الاختلافات :

$$\sum(Y - Y^{\backslash})^2 = \sum e^2 = e_1^2 + e_2^2 .....$$

فإذا كان حاصل مجموع مربع الاختلافات صغيراً أمكننا حينئذٍ اعتبار أن خط العلاقات يوفق بين النقاط بصورة جيدة. وتعرف الطريقة التي يمكننا بها اختيار الثوابت b_0, b_1 بحيث نوفق خط العلاقة $Y^{\backslash} = b_0 + b_1 X$ للبيانات الفعلية، وبشكل يجعل مجموع مربع البواقي (The sum of squares of residuals) نهاية صغرى (Minimum - Minimum) بطريقة المربعات الصغرى. ويتم ذلك بإيجاد الحل للمعادلات الطبيعية (The normal equations – Les équations normales) الناتجة عن التفاضل الجزئي (Partial derivative – Dérivée partielle) لمجموع مربع البواقي بالنسبة للثوابت b_0 و b_1 وجعل المفاضلة الجزئية مساوية للصفر كالآتي :

$$Y = b_0 + b_1 X + e$$

$$e = Y - b_0 - b_1 X$$

$$\sum e^2 = \sum(Y - b_0 - b_1 X)^2$$

$$\frac{\partial \sum e^2}{\partial b_0} = 2\sum(Y - b_0 - b_1 X)(-1) = 0$$

$$\sum Y = b_0 n + b_1 \sum X \qquad (1)$$

$$\frac{\partial \sum e^2}{\partial b_1} = 2\sum\left(Y - b_0 - b_1 X\right)(-X) = 0$$

$$\sum XY = b_0 \sum X + b_1 \sum X^2 \qquad (2)$$

تدعى المعادلتين (1) و (2) بالمعادلات الطبيعية. علماً أنه كان من الممكن الحصول عليهما بسهولة ودون استخدام التفاضل وذلك باستخدام معادلة الانحدار:

$$Y = b_0 + b_1 X$$

وبجمع طرفي معادلة الانحدار نحصل على:

$$\sum Y = b_0 n + b_1 \sum X$$

وبضرب طرفي معادلة الانحدار بالمتغير X وبالجمع نحصل على:

$$\sum XY = b_0 \sum X + b_1 \sum X^2$$

أما الثابت b_0 فيتم الحصول عليه من قسمة طرفي المعادلة (1) على n:

$$\frac{\sum Y}{n} = \frac{b_0\, n}{n} + b_1 \frac{\sum X}{n}$$

$$\overline{Y} = b_0 + b_1 \overline{X}$$

$$b_0 = \overline{Y} - b_1 \overline{X} \qquad (3)$$

أما معامل الانحدار b_1 فيمكن الحصول عليه من المعادلتين (1) و(2) وذلك باستخدام قاعدة كرايمر (Cramer's rule – Règle de Cramer) في حل معادلات متجانسة في متغيرين اثنين كالآتي :

$$b_1 = \frac{\begin{vmatrix} n & \sum Y \\ \sum X & \sum XY \end{vmatrix}}{\begin{vmatrix} n & \sum X \\ \sum X & \sum X^2 \end{vmatrix}} = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2}$$

$$b_1 = \frac{\sum XY}{\sum X^2} \qquad (4)$$

حيث أن $\sum XY$ تمثل مجموع حواصل ضرب انحرافات القيم عن الوسط الحسابي للمتغيرين X وY. كما وأن $\sum X^2$ تمثل مجموع مربع الانحرافات لقيم المتغير X عن وسطه الحسابي. علماً أن

تعرف الثوابت b_0 و b_1 بمعاملات الانحدار ، وتجدر الإشارة هنا إلى ضرورة التمييز بين:

- معامل الانحدار البسيط ومعامل الانحدار الجزئي.
- معامل الانحدار بالوحدات الخام ومعامل الانحدار بالوحدات المعيارية *.
- معامل الانحدار للمجتمع الإحصائي ومعامل الانحدار التقديري من العينة.

بما أن معادلة التوقع $\hat{Y} = b_0 + b_1 X$ هي معادلة الخط المستقيم لذلك فإن معاملات الانحدار تحدد موقع خط الانحدار حيث تمثل b_0 البعد بين نقطة تقاطع خط الانحدار مع الإحداثي العمودي Y، ونقطة الأصل (The origin – L'origine) **. أما الثابت b_1 فيمثل ميل المستقيم (The slope – La pente) ويقيس معدل التغير في المتغير التابع والمرتبط بتغير وحدة قياس واحدة في المتغير المستقل. وتعتبر b_0 و b_1تقديرات من بيانات العينة لمعالم الانحدار β وα في المجتمع الإحصائي. علماً أنه إذا تضمنت معادلة الانحدار على متغير تابع ومتغير واحد فإن b_1 تعرف بمعامل الانحدار البسيط (Simple regression coefficient – Coefficient de la régression simple) أما إذا تضمنت معادلة الانحدار على أكثر من متغير مستقل، فحينئذ يعرف كل من معاملات الانحدار b_iبمعامل الانحدار الجزئي (Partial regression coefficient – coefficient de la régression partielle) والذي يقيس معدل التغير في Y نتيجة تغير المتغير المستقل X بوحدة قياس واحدة، مع بقاء أثر بقية المتغيرات المستقلة الأخرى ثابتاً (keeping the effect of other independent variables constant – En gardant l'effet des autres variables indépendantes constant). إي أن معامل الانحدار الجزئي b_i يستخدم في الانحدار المتعدد (Multiple regression – Régression multiple) ويقيس العلاقة بين ما تبقى من المتغير ما تبقى من المتغير المستقل (بعد حذف أثر بقية المتغيرات المستقلة منه) وبين المتغير التابع.

تجدر الإشارة إلى أننا كثيراً ما نضطر إلى استخدام متغيرات اقتصادية مقاسة بوحدات قياس مختلفة كأن يقاس السماد بالباوند، وتقاس الأمطار بالإنش، في حين يقاس الحبوب بمكيال مثل البوشل(Bushels)، وبالتالي يصعب على الباحث في مثل هذه الحالة مقارنة معاملات الانحدار

* لاحظ أننا لو استخدمنا معادلة الانحدار في شكل انحرافات عن الوسط الحسابي (In deviation form – Déviation de la moyenne) فإنّ الثابت b_0 عندئذٍ يساوي صفراً في معادلة الانحدار $y = b_1 x + e$. ويتم الحصول على b_1 بالمفاضلة الجزئية لمجموع مربع البواقي بالنسبة للمعامل b_1 كالآتي:

$$\frac{\partial \sum e^2}{\partial b_1} = 2\sum\left(y - bx\right)\left(-x\right) = 0$$

$$\sum xy = b_1 \sum x^2$$

$$b_1 = \frac{\sum xy}{\sum x^2}$$

* يقصد بالوحدات الخام (Raw scores – Scores brutes) الوحدات التي لم تخضع بعد لأية معالجة إحصائية.

** لاحظ أن أفضل طريقة لرسم خط الانحدار هو أن نرسم خط مستقيم يصل بين نقطة التقاطع (The intercept – La constante) b_0 والنقطة المؤلفة من الإحداثيات $\left(\bar{X}, \bar{Y}\right)$ ، وذلك لأنه عند $\left(X = \bar{X}\right)$ فإن $\hat{Y} = \bar{Y}$ ، ويمكن توضيح ذلك كالآتي:

$$\hat{Y} = b_0 + b_1 \bar{X}$$

$$\hat{Y} = \bar{Y} - b_1 \bar{X} + b_1 \bar{X} = \bar{Y}$$

b_1 , b_2

β

$$Z_x = \frac{X - \overline{X}}{S_X}$$

β_0

$$Z_Y = \beta Z_X$$

حيث تمثل β معامل الانحدار البسيط بالوحدات المعيارية ،علماً أن:

$$\beta = \frac{\sum Z_X Z_Y}{\sum Z_X^2} = \frac{\sum Z_X Z_Y}{n-1} = r$$

وعلى الرغم من أن معامل الانحدار البسيط بالوحدات المعيارية $\boldsymbol{\beta}$ ، يساوي معامل الارتباط البسيط $\boldsymbol{r}$ ، إلا أن معامل الانحدار البسيط بالوحدات المعيارية لن يساوي معامل الانحدار البسيط بالوحدات الخام إلا إذا تساوت الانحرافات المعيارية للمتغيرين X و Y وذلك أنّ*:

$$\boldsymbol{b = \beta \frac{S_Y}{S_X} = r \frac{S_Y}{S_X}}$$

*** يجب التنويه إلى ضرورة الانتباه إلى أن معامل الانحدار بالوحدات المعيارية β هو أيضاً تقدير من بيانات العيّنة لمعامل الانحدار في المجتمع الإحصائي $\boldsymbol{\beta}$, حيث يستخدم الرمز β في الحالتين.

**** لاحظ أنه عند استخدام الوحدات المعيارية فإن كل قيمة في المتغير تكون مقاسة في شكل انحراف عن الوسط الحسابي وبالتالي فإن خط الانحدار سيمر من نقطة الأصل حيث أن $Z_{\overline{X}} = Z_{\overline{Y}} = 0$ وبالتالي فإن $\beta_0 = Z_{\overline{Y}} - \beta_1 Z_{\overline{X}} = 0$

* لاحظ أن الوسط الحسابي للقيم المتوقعة $\hat{Y}$ يساوي الوسط الحسابي للقيم الفعلية Y لأن $\sum e = 0$ دائماً، كما وأن $\sum \hat{Y} = \sum (Y + e) = \sum Y$ ، وبالتالي فإن $\overline{Y} = \overline{\hat{Y}}$. أما تباين القيم المتوقعة $S_{\hat{Y}^2}$ فيساوي:

$$S_{\hat{Y}^2} = \frac{1}{n}\sum\left(\hat{Y} - \overline{Y}\right)^2 = \frac{1}{n}\sum\left(b_0 + b_1 X - \overline{Y}\right)^2$$

$$S_{\hat{Y}}^{\;2} = \frac{1}{n}\sum\left(\overline{Y} - b_1\overline{X} + b_1 X - \overline{Y}\right)^2 = \frac{1}{n}\sum\left[b_1\left(X - \overline{X}\right)\right]^2$$

$$S_{\hat{Y}}^{\;2} = b_1^2 \frac{\sum\left(X - \overline{X}\right)^2}{n} = b_1^2 S_X^2$$

وبقسمة طرفي المعادلة على التباين الكلي للمتغير التابع S_y^2 نحصل على:

$$r^2 = \frac{S_{\hat{Y}}^{\;2}}{S_y^2} = b_1^2 \frac{S_X^2}{S_y^2}$$

انظر في ذلك Aigner p:29

$$\beta = r = b\frac{S_X}{S_Y}$$

وتقيس β معدل التغير في Y مقاساً بالوحدات المعيارية فيما لو تغير X بانحراف معياري واحد. فلو فرضنا أن الانحراف المعياري لمتغير مستقل(كالأمطار مثلاً) يساوي ثلاث إنشات $S_X = 3$. وأن الانحراف المعياري لمتغير تابع كإنتاج الحبوب مثلاً (يساوي أربع مكايل $S_Y = 4$). ولو فرضنا أننا حصلنا على $\beta = 0.60$ ، فحينئذ يمكننا القول أنه لو تغيّرت الأمطار بالانحراف معيراي واحد فإنّ إنتاج الحبوب سيتغيّر بمقدار 0.60 من الانحراف المعياري في Y، بمعنى أنّه لو تغيّر سقوط الأمطار بمقدار 3 إنشات، فإنّ إنتاج الحبوب سيتغيّر بمقدار $0.60 \times 4 = 2.40$ من المكاييل. علماً أنّه إذا تضمّنت معادلة الانحدار على أكثر من متغيّر مستقل، فإنّ معاملات الانحدار بالوحدات المعيارية، تُعرف عندئذ باسم معاملات الانحدار الجزئيّة بالوحدات المعيارية (Standard Partial regression coefficients – Coefficients standardisés de la régression partielle)، ويقيس كل منها معدل التغيّر في Y نتيجة تغير X بانحراف معياري واحد، مع بقاء أثر بقية المتغيرات المستقلة الأخرى ثابتاً.

وتجدر الإشارة أخيراً، إلى أنه يمكنّنا كتابة المعادلات الطبيعية في شكل مصفوفات(Matrices - Matrices) كالآتي:

$$Y = Xb + e$$

حيث يمكن الحصول على معاملات الانحدار باستخدام المصفوفات كالآتي:

$$b = \left(X'X\right)^{-1}\left(X'Y\right)$$

علماً أنّ $\left(X'X\right)^{-1}$ هي مقلوب مصفوفة التبايّن والتغايّر للموجه X.[3]

الخصائص الإحصائية لتقديرات المربعات الصغرى:

تتميّز التقديرات التي نحصل عليها من بيانات العينة باستخدام طريقة المربعات الصغرى بأنّها أكفأ (Best - Meilleur) تقديرات خطية (Linear - Linéaire) غير متحيزة (Unbiased – Non biaisée) للمعالم α ,β في المجتمع الإحصائي. ويعود السبب في ذلك إلى أنّ b هي تركيب خطي (Linear combination – Combinaison linéaire) للقيم الفعلية للمتغيّر التابع Y فمن المعلوم أنّ *

$$b = \frac{\sum XY}{\sum X^2}$$

---

[3] Drapper & Smith ,. "Applied regression Analysis" john Wiley & Sons, inc. 1966. pp:44-52.
وتجدر الإشارة إلى أننا سوف نتناول كيفية صياغة المعادلات الطبيعية في شكل مصفوفات وبالتفصيل في الفصل الثالث ص ص 109-111.

* سيتم التركيز على الخصائص الإحصائية للمعامل b_1 دون b_0 ، لأن b_1 أكثر أهمية في الاقتصاد القياسي من b_0 .

ونظراً لأنّ المتغيّر المستقل X هو متغيّر ثابت في المجتمعات الفرعية لذلك فإنه يمكن اعتبار القيمة $\frac{X}{\sum X^2}$ مساوية إلى التثقيل (Weight- Pondération)، والذي يُرمز له بالحرف ω فتصبح b مجموعاً مرجحاً (Weighted sum – Somme pondérée) لقيم المتغيّر التابع.[4]

$$b = \frac{\sum xy}{\sum x^2}$$
$$b = \sum \omega y$$
$$b = \sum \omega (Y - \overline{Y})$$
$$b = \sum \omega Y - \overline{Y} \sum \omega$$
$$b = \sum \omega Y - \overline{Y} \frac{\sum x}{\sum x^2}$$

ونظراً لأنّ $\sum x = 0$، إذاً

$$b = \sum \omega Y$$

بمعنى أنّ قيم Yمستقلة من الناحية الإحصائية (Statistically independent – statistiquement indépendante)، فلو كانت Y_1 كبيرة فإنّ Y_2 قد تكون كبيرة أو صغيرة أو صفرا، وفي حالة سحب عيّنات معادة (Repeated random sample – échantillon aléatoire répété) من نفس المجتمع الإحصائي للمتغيّر Y حيث تكون X ثابتة فإنّ قيم Yتكون مستقلة وبالتالي فإنّ:

$$b = \omega_1 Y_1 + \omega_2 Y_2 + \omega_3 Y_3 + \ldots.. + \omega_n Y_n$$
$$b = \sum \omega Y$$
$$E(b) = E\left(\sum \omega Y\right)$$
$$E(b) = \sum \omega \, E(Y)$$
$$E(b) = \sum \omega (\alpha + \beta X)$$
$$E(b) = \alpha \sum \omega + \beta \sum \omega X$$
$$E(b) = \beta^{5}$$

---

[4] انظر: Wonnacott & wonnacott pp: 19-20

[5] لاحظ أن:

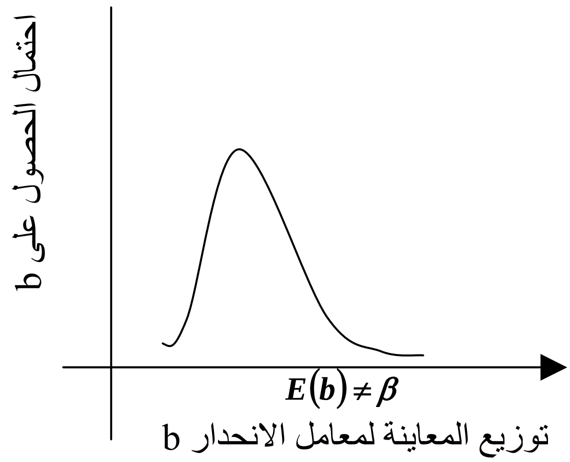
$$\sum \omega = \sum \frac{X}{\sum X^2} = \frac{\sum X}{\sum X^2} = 0$$
$$\sum \omega X = \sum \frac{X}{\sum X^2} . X = \frac{\sum (X - \overline{X}) . X}{\sum X^2}$$
$$\sum \omega X = \frac{\sum X^2 - \overline{X} \Sigma X}{\Sigma X^2} = 1$$

لاحظ أيضاً أنه كان من الممكن إثبات أن b هي تقدير غير متحيز لقيمة β كالآتي:

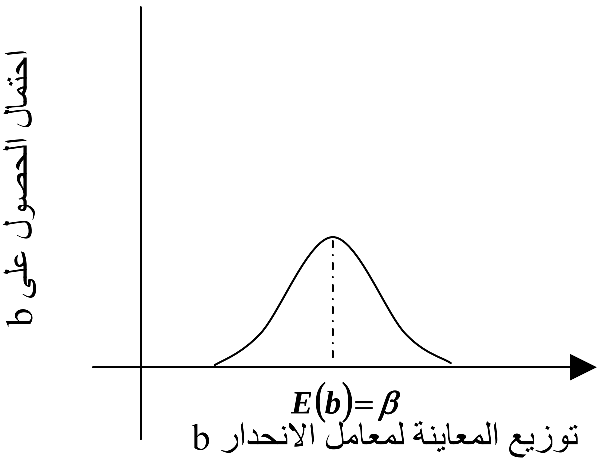
$$b = \frac{\Sigma xy}{\Sigma x^2} = \frac{\Sigma x(\beta x + e)}{\Sigma x^2}$$
$$b = \frac{\beta \Sigma x^2}{\Sigma x^2} + \frac{\Sigma xe}{\Sigma x^2}$$
$$E(b) = \beta$$

بمعنى أنّ ***b*** هي تقدير غير متحيّز (Unbiased estimate – estimation non biaisée) لقيمة معامل الانحدار ***β*** في المجتمع الإحصائي، بمعنى أنّ القيمة المتوقعة (Expected value) لتوزيع المعاينة (Sampling distribution - échantillonnage) للمعامل ***b*** يساوي إلى قيمة ***β*** في المجتمع الإحصائي. ويمكن توضيح ذلك بيانياً كالآتي:

تقدير متحيز



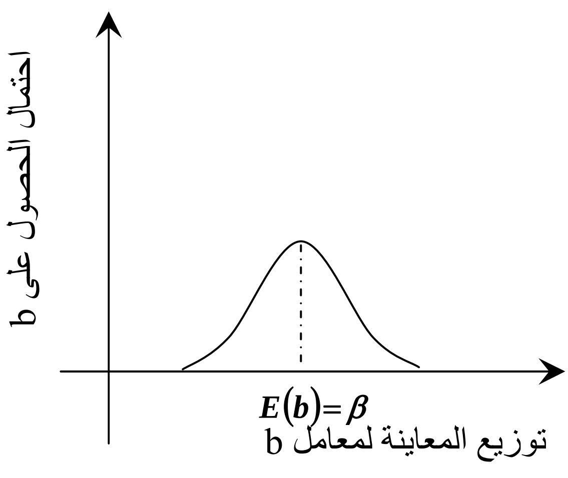
تقدير غير متحيز



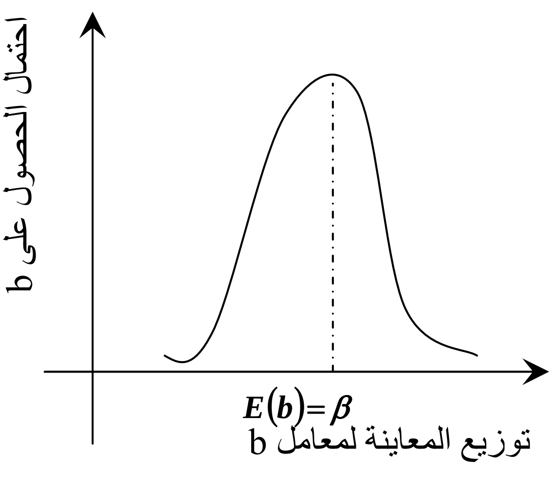
علماً أنّه، عندما يُقال أنّ ***b*** هي تقدير غير متحيّز لقيمة ***β***، فهذا لا يعني ان ***b = β***، في حل عينه عشوائية مسحوبة من نفس المجتمع الإحصائي، لكن عند سحب عيّنات عشوائية معادة من نفس المجتمع الإحصائي، فإن الوسط الحسابي لتوزيع المعاينة للمعامل ***b*** سوف يساوي إلى قيمة المعامل الحقيقي في المجتمع الإحصائي ***β***.

وتتميّز التقديرات لطريقة المربعات الصغرى بأنها أكثر كفاءة (Best unbiased = efficient – meilleur non biaisé = efficient) من بين كل التقديرات غير المتحيّزة*. ويمكن توضيح ذلك بيانياً كالآتي:

تقدير غير كفوء



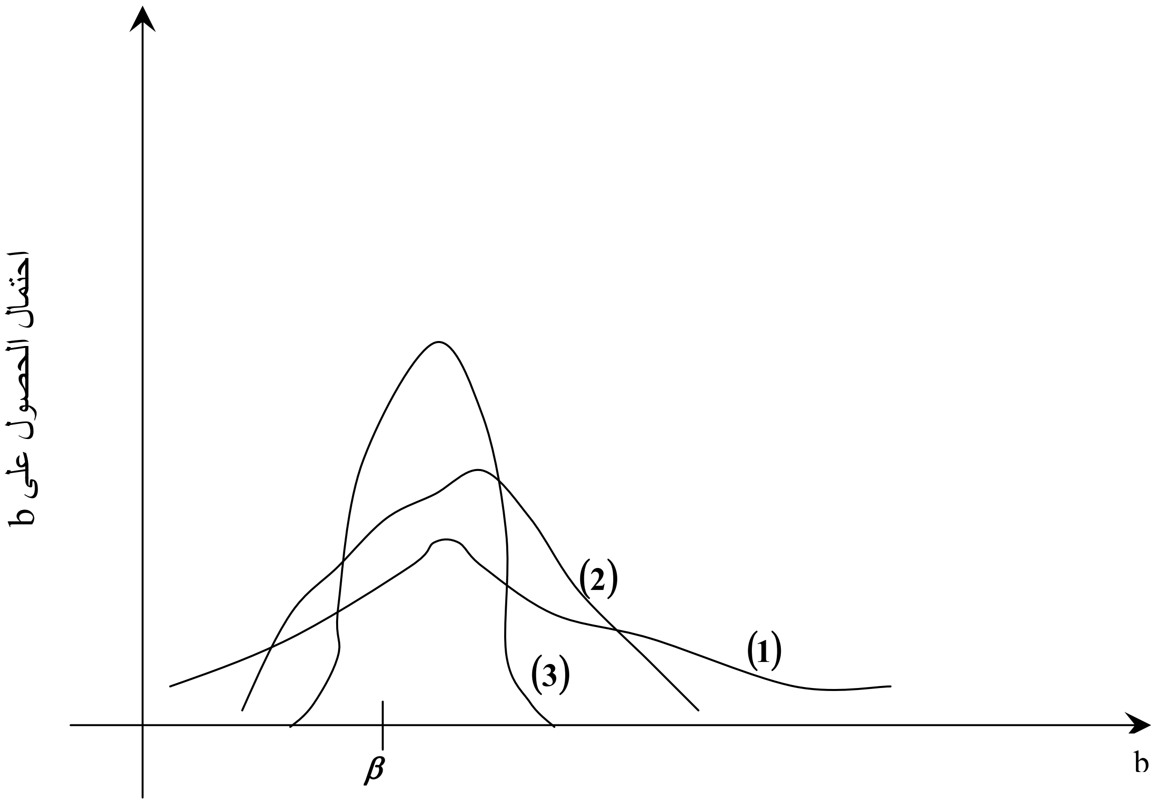
تقدير كفوء



* لاحظ مثلاً: انه يقال في الإحصاء بأن الوسط الحسابي (Arithmetic mean) أكثر كفاءة من الوسيط (Median)، وذلك على الرغم من أن القيمة المتوقعة لتوزيع المعاينة، للوسط الحسابي أو للوسيط، تساوي قيمة الوسط الحسابي في المجتمع الإحصائي، وذلك لأن تباين توزيع المعاينة للأوساط الحسابية يكون أقل من تباين توزيع المعاينة للوسيط.

ويتميّز التقدير الكفوء بأنه أقل تبايناً (Variance - Variance) من التقدير غير الكفوء، وهذا يعني أنّ فترة الثقة ستكون أصغر (Smaller confidence interval – le plus petit intervalle de confiance)، وبالتالي يوجد احتمال أكبر للحصول على نتائج جوهرية من الناحية الإحصائية (Statistically significant – Statistiquement significatif) وباختصار يمكننا القول أنّه إذا حصلنا على التقديرات غير المتحيّزة $b_0 + b_1 X$ بطريقة المربعات الصغرى فإنه عند سحب عيّنات معادة من نفس المجتمع الإحصائي ، فإن تشتت خطوط الانحدار $b_0 + b_1 X$ حول خط الانحدار الحقيقي $\alpha + \beta X$، سيكون أقل من تشتت خطوط الانحدار غير المتحيّزة $c + dx$ والتي تمّ الحصول عليها بطريقة أخرى غير طريقة المربعات الصغرى.

وتجدر الإشارة أخيراً إلى أنّ تقديرات المربعات الصغرى تتميّز بكونها تقترب من القيمة الحقيقية β في المجتمع الإحصائي بازدياد حجم العينة (Consistent estimator – Estimateur consistant). فكلّما زادة حجم العيّنة كلما اقتربت قيمة b من قيمة β (Asymptotic unbiased estimator –Estimateur asymptotiquement non biaisé) كما يتوضح في الرسم البياني الآتي:



نلاحظ في الرسم البياني أعلاه أنّه تمّ الحصول على توزيع (1) عندما كانت n صغيرة، ثمّ عندما أصبحت n كبيرة حصلنا على التوزيع (2) وعندما أصبحت n كبيرة جداً حصلنا عل التوزيع (3)، ويمكننا القول أنه كلما أصبح حجم العيّنة كبيراً، كلّما اقتربت قيمة b من قيمة β.

دعنا نفترض للتبسيط أن أحد الباحثين يرغب في دراسة علاقة استهلاك الإسمنت بالناتج القومي، لبعض الدّول العربية فجمع البيانات الموضحة في الجدول رقم (2):

الجدول رقم(2)

الناتج القومي الإجمالي بسعر السوق (مليون دولار)،

واستهلاك الإسمنت (ألف طن) لبعض الدول العربية عام 1978[6]

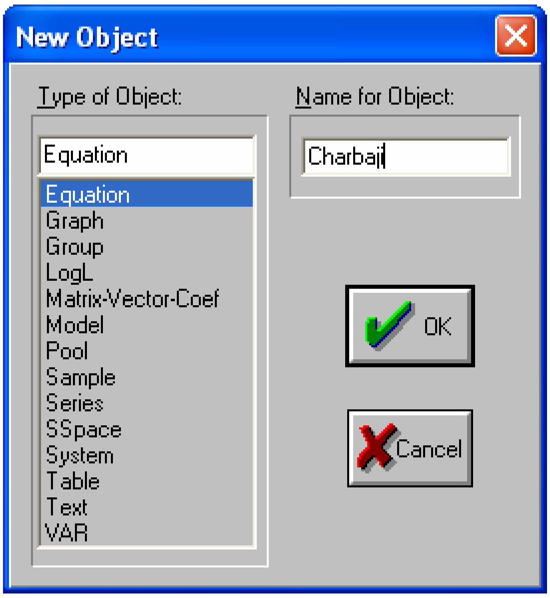
النــاتج القومي الإجمالي X	استهلاك الاسمنت Y
22567.6	4021
23124.4	6196
13778.2	2394
15281.3	2376
65815.6	8962
19045.7	3182
1877.1	598
5963.5	1672
8277.3	3000
24715.2	3808
1856.4	1192
12427.1	3504
6458.6	435
620.3	65
2618.5	136

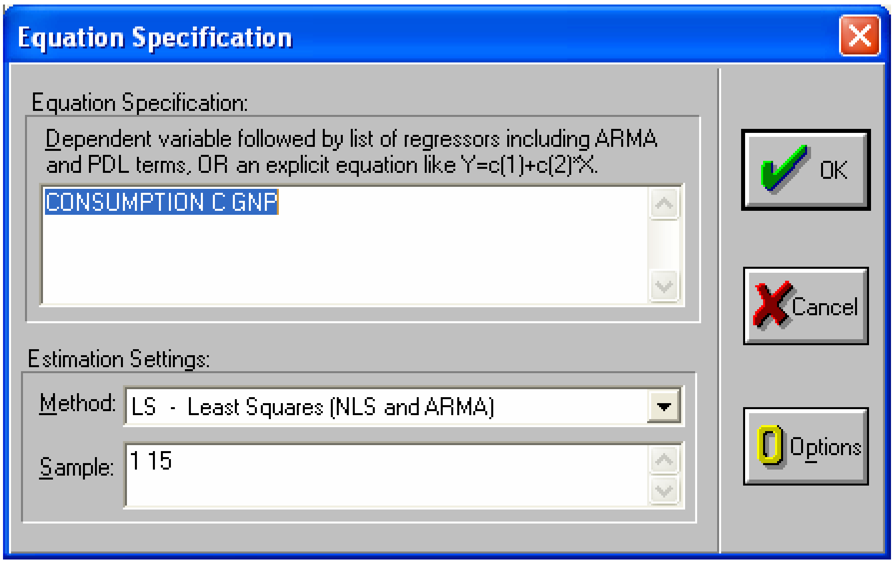
$$Y^{\backslash} = 718.07 + 0.137\,X$$

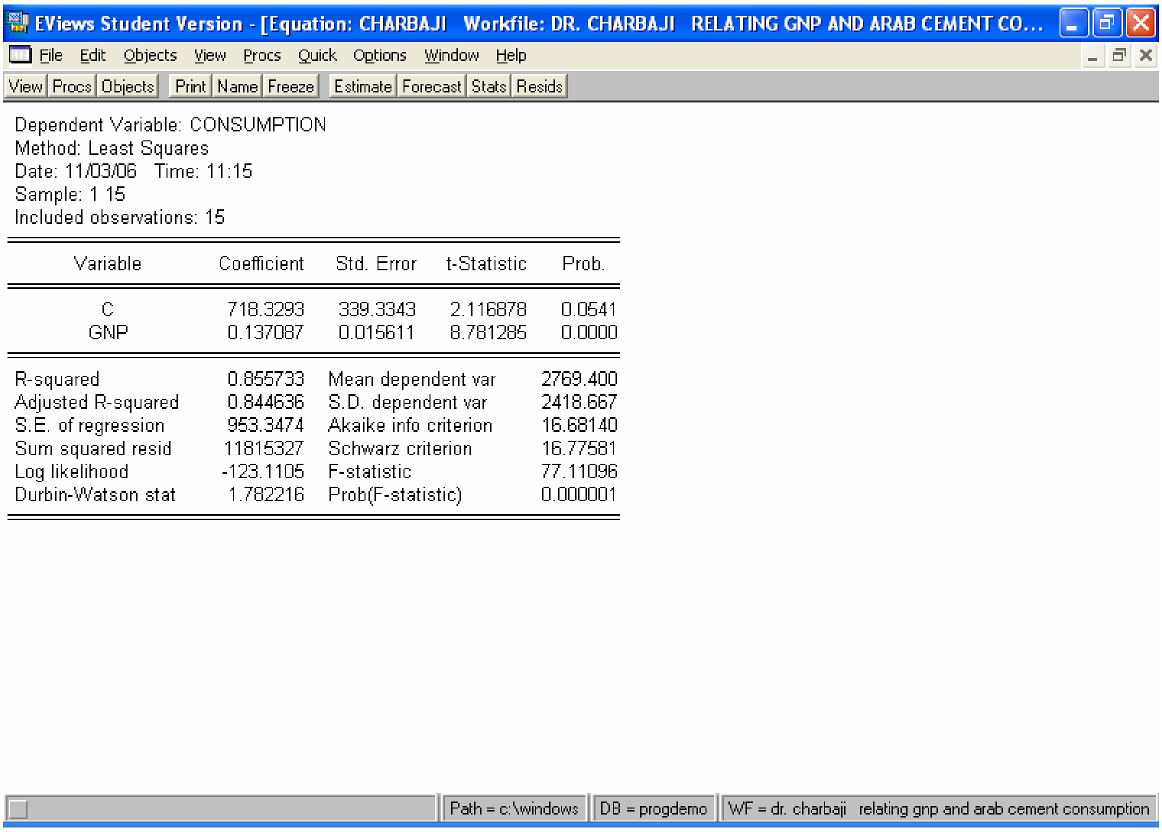
$$r^2 = 0.86$$

[6] " التقرير الاقتصادي العربي الموحد" الصادر عن: الأمانة العامة لجامعة الدول العربية، صندوق النقد العربي، الصندوق العربي للإنماء الاقتصادي والاجتماعي عام 1981 ص 190 و 208

OBJECT / NEW OBJECT / EQUATION







EViews Student Version - [Equation: CHARBAJI Workfile: DR. CHARBAJI RELATING GNP AND ARAB CEMENT CO...

File Edit Objects View Procs Quick Options Window Help

View Procs Objects Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: CONSUMPTION
Method: Least Squares
Date: 11/03/06 Time: 11:15
Sample: 1 15
Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	718.3293	339.3343	2.116878	0.0541
GNP	0.137087	0.015611	8.781285	0.0000

R-squared	0.855733	Mean dependent var	2769.400
Adjusted R-squared	0.844636	S.D. dependent var	2418.667
S.E. of regression	953.3474	Akaike info criterion	16.68140
Sum squared resid	11815327	Schwarz criterion	16.77581
Log likelihood	-123.1105	F-statistic	77.11096
Durbin-Watson stat	1.782216	Prob(F-statistic)	0.000001

Path = c:\windows | DB = progdemo | WF = dr. charbaji relating gnp and arab cement consumption

نلاحظ في الجدول أعلاه أنّ 86% من الاختلافات الكلية لاستهلاك الإسمنت في الدول العربية تمّ تحديدها (تفسيرها) عن طريق معرفتنا بالاختلافات الكلية في الناتج القومي الإجمالي لهذه الدّول. أمّا فيما يتعلّق بمعاملات الانحدار فيجب الانتباه إلى أنّ البيانات أعلاه مأخوذة في نقطة زمنية محددة (cross-section data - données en coupes instantanées)، فهي ليست بيانات لسلاسل زمنية (time series data – données chronologiques)، ويمكن تفسير معامل الانحدار $\beta = \mathbf{0.137}$ على أنّه إذا زاد الناتج القومي الإجمالي لإحدى الدّول العربية على دولة أخرى، بوحدة قياس واحدة، فإنّ استهلاكها المتوَقّع من الإسمنت، سيزيد عن تلك الدّولة بمقدار 0.137 من وحدة القياس، آخذين في الاعتبار أنّ وحدات القياس المتغيرين $\boldsymbol{X}$ و $\boldsymbol{Y}$ مختلفة، حيث أنّ وحدة القياس للمتغيّر $\boldsymbol{Y}$ تساوي ألف طن، في حين أنّ وحدة القياس للمتغير $\boldsymbol{X}$ تساوي مليون دولار، وبالتالي يمكننا القول، أنّ إذا زاد الناتج الإجمالي لدولة عربية على دولة أخرى بمقدار مليون دولار، فإنّ الاستهلاك المتوقع من الإسمنت لهذه الدولة، سيزيد على الدولة الأخرى بمقدار 137 طن. أمّا فيما يتعلّق بنقطة التقاطع $\alpha = \mathbf{718.07}$، فتفسر على أنها الاستهلاك المتوّقع من الإسمنت بالألف طن عندما يكون الناتج القومي الإجمالي مساوياً للصفر $\boldsymbol{X} = \mathbf{0}$، علماً بأنّ α لا تفيد كثيراً في التفسير وإنما تفيدنا في تحديد موقع خط الانحدار.[7] وتجدر الإشارة هنا إلى أنّنا استعملنا α و β بدلاً من b_0 و b_1 في المثال أعلاه بهدف الشرح والتبسيط، حيث اعتبرنا أنّ البيانات في الجدول رقم (2) تمثّل بيانات المجتمع الإحصائي بأكمله.

[7] انظر في نهاية هذا الفصل إلى الطرق المختلفة التفصيلية للحصول على معاملات الانحدار والارتباط والتحديد. كما وتجدر الإشارة إلى أنه لو كانت هذه البيانات لسلاسل زمنية (Time series data) فحينئذٍ تفسر β على أنها معدل التغير المتوقع في Y نتيجة تغير X بوحدة قياس واحدة (بفترة زمنية واحدة).

الجدير ذكره أن زيادة حجم العيّنة يؤدي إلى اقتراب قيمة معامل الانحدار للعيّنة b من قيمة معامل الانحدار في المجتمع الإحصائي β ممّا يدل على أن زيادة حجم العينة يجعل منحنيات الانحدار للعينات $(b_0 + b_1 X)$ قريبة من منحنى الانحدار $(\alpha + \beta X)$ للمجتمع الإحصائي[8].

يتضح مما سبق ذكره، أنه كلما كبر حجم العينة كلما اقتربت خطوط الانحدار للعينات من خط الانحدار الحقيقي للمجتمع الإحصائي. ويثار التساؤل التالي: هل يوجد مقياس محدد يقيس تشتت قيم b حول β مهما كان حجم العينة؟ لا شك أن أفضل مقياس لتشتت قيم b حول β هو تباين معامل الانحدار b والذي يرمز له بالرمز σ_b^2 فإذا كان σ_b^2 صغيراً أمكننا القول أننا حققنا جودة في توفيق منحنى الانحدار والعكس صحيح[9].

معامل التحديد (Coefficient of determination – Coefficient de détermination)

لا شك أن موضوعي الارتباط والانحدار هما موضوعان متشابهان للغاية فعندما يريد الباحث دراسة انحدار المتغير التابع Y (الكمية المعروضة مثلاً) على المتغير المستقل X (السعر مثلاً)، فإنه في الحقيقة يعني كيفية تفسير (شرح) التباين (variance - variance) أو الاختلافات (variation - variation) في المتغير التابع عن طريق معرفته بالاختلافات أو التباين في المتغير المستقل[10].

[8] تجدر الإشارة إلى أن فكرة الانحدار (Regression) تعود تاريخياً إلى العالم غالتون (Galton). فقد لاحظ غالتون في بحوثه عن الوارثة (Inheritance) أنه لو افترضنا أن طوال القامة يتزوجون من طوال القامة مثلهم، وأن قصار القامة سيتزوجون قصارا مثلهم، فحينئذٍ سيلد طوال القامة، طوالاً مثلهم وسيلد قصار القامة قصاراً مثلهم، وينقسم المجتمع عندئذٍ إلى مجموعتين، بحيث تتضمن المجموعة الأولى على العمالقة، في حين تتضمن المجموعة الثانية الأقزام. لكن غالتون لاحظ أنه عبر الأجيال، فإن الأحفاد من المجموعتين تنحدر في طول قامتها نحو الوسط الحسابي للطول في المجتمع الإحصائي. ومن ظاهرة اتجاه الطول نحو الوسط الحسابي (Regression toward mediocrity) اشتقت كلمة الانحدار. انظر:

Fred N Kerlinger and Elazar j. pedhazur., Multiple Regression in Behavioral Research. Holt, Rinhert and Winston, lnc., 1973,p:18

[9] انظر فترة الثقة ص:67

[10] التباين لأي متغير هو عبارة عن حاصل قسمة الاختلافات الكلية على درجات الحرية (degree of freedom)، ومثال ذلك تباين المتغير التابع Y:

$$\sigma_y^2 = \frac{\Sigma\left(Y - \overline{Y}\right)^2}{n-1}$$

إحصائياً، لو أردنا التنبؤ بقيم المتغير التابع Y من المتغير المستقل X، فإن القدرة على التنبؤ تتوقف على معامل الارتباط بين المتغيرين X و Y. فإذا كانت قيمة الارتباط بين المتغيرين صفراً، فحينئذٍ يمكن لكل قيمة من قيم المتغير المستقل، أن تهدينا للتنبؤ إلى الوسط الحسابي (The mean – La moyenne) من المتغير التابع Y حيث تصبح $\bar{Y}$ أفضل قيمة ممكن التنبؤ بها في هذه الحالة. أما إذا كانت قيمة معامل الارتباط بين الصفر $r = 0.0$ وبين الواحد الصحيح $r = \pm 1.0$ فحينئذٍ تزيد القدرة على التنبؤ بشكلٍ أفضل مما كان عليه عندما $r = 0.0$. أما إذا كان معامل الارتباط مساوياً للواحد الصحيح $r = \pm 1.0$، فحينئذٍ تصبح لدى الباحث قدرة تامة على التنبؤ بقيم المتغير التابع Y من قيم المتغير المستقل X بدون أخطاء[11].

يعرف مجموع مربع الانحرافات الكلية (The sum of squares – La somme des carrés) لقيم المتغير التابع Y عن الوسط الحسابي $\bar{Y}$ بالاختلافات الكلية (Total variation – Variation totale) في المتغير التابع، الذي يراد تفسيره عن طريق معرفتنا بالاختلافات الكلية لمتغير أو أكثر من المتغيرات المستقلة، ويرمز له بالرمز SS_t. ويتكون الاختلاف الكلي من اختلاف مفسر (Explained variation – Variation expliquée) ومن اختلاف غير مفسر (Unexplained variation – Variation inexpliquée)، أما الاختلاف المفسر فهو ذلك الجزء من الاختلاف الكلي في المتغير التابع Y والذي يتم تحديده بمعادلة الانحدار، ويرمز له بالرمز $SS_{\text{Re}\,g}$ ويقصد به مجموع مربع الانحرافات في المتغير التابع Y والمحددة بالانحدار (Sum of squares of Y due regression – Somme des carrés de la regression Y). أما الاختلاف غير المفسر فهو ذلك الجزء من الاختلاف الكلي في المتغير التابع والذي لم نتمكن من تحديده عن طريق العلاقة بين المتغير التابع Y والمتغير المستقل X.

(The variation in Y which remains unexplained by the estimated relationship between X and Y – La variabilité dans Y qui n'est pas prise en compte par la régression)

ويعرف الاختلاف غير المفسر بالبواقي (Residuals) ويرمز له بالرمز $SS_{\text{Re}\,sd}$. فمن المعلوم أن:

$$Y = b_0 + bX + e$$

$$Y' = b_0 + bX$$

$$Y' = \bar{Y} - b\bar{X} + bX$$

$$Y' = \bar{Y} + b\left(X - \bar{X}\right) \qquad (1)$$

وبطرح طرفي المعادلة (1) من Y نحصل على:

$$Y - Y' = Y - \bar{Y} - b\left(X - \bar{X}\right) \qquad (2)$$

وبتربيع طرفي المعادلة (2) والجمع نحصل على:

[11] يقصد بالتنبؤ (Prediction - Prévision) استخدام تحليل الانحدار (Regression analysis – Analyse de régression) في توقع (تقدير) القيمة المتوقعة للمتغير التابع $E\left(Y/X\right)$ من خلال معرفتنا بقيمة متغير أو أكثر من المتغيرات المستقلة. ويعتبر التنبؤ جزءاً من تفسير الظواهر (Explanation - Explication)، ويختلف التنبؤ عن التكهن (Forecasting - Prévision) في أن التنبؤ هو تعبير شرطي (conditional - Conditionnelle) بمعنى أنه إذا كان لدينا قيمة معينة للمتغير X فحينئذٍ يمكننا توقع $\hat{Y}$ باستخدام معادلة الانحدار.

$$\sum\left(Y-Y'\right)^2=\sum\left[\left(Y-\overline{Y}\right)-b\left(X-\overline{X}\right)\right]^2$$
$$\sum\left(Y-Y'\right)^2=\sum\left(Y-\overline{Y}\right)^2+b^2\sum\left(X-\overline{X}\right)^2-2b\sum\left(Y-\overline{Y}\right)\left(X-\overline{X}\right) \quad (3)$$

علماً أن الحد الأخير من المعادلة (3) يساوي :

$$-2b\sum\left(Y-\overline{Y}\right)\left(X-\overline{X}\right)=-2b.b\sum\left(X-\overline{X}\right)^2$$

ذلك لان:

$$b=\frac{\sum\left(Y-\overline{Y}\right)\left(X-\overline{X}\right)}{\sum\left(X-\overline{X}\right)^2}$$

وباستبدال الحد الأخير في المعادلة (3) بقيمته نحصل على:

$$\sum\left(Y-Y'\right)^2=\sum\left(Y-\overline{Y}\right)^2+b^2\sum\left(X-\overline{X}\right)^2-2b^2\sum\left(X-\overline{X}\right)^2$$
$$\sum\left(Y-Y'\right)^2=\sum\left(Y-\overline{Y}\right)^2-b^2\sum\left(X-\overline{X}\right)^2$$

(4)

لكننا نعلم من المعادلة (1) أن:

$$b^2\sum\left(X-\overline{X}\right)^2=\sum\left(Y'-\overline{Y}\right)^2$$

وباستبدال هذه القيمة في المعادلة (4) نحصل على:

$$\sum\left(Y-Y'\right)^2=\sum\left(Y-\overline{Y}\right)^2-\sum\left(Y'-\overline{Y}\right)^2$$

إذن:

(5)

$$\sum\left(Y-\overline{Y}\right)^2=\sum\left(Y'-\overline{Y}\right)^2+\sum\left(Y-Y'\right)^2$$
$$\downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow$$

الاختلافات غير المفسرة + الاختلافات المفسرة = الاختلافات الكلية في المتغير التابع

وبقسمة طرفي المعادلة (5) على مجموع مربع الاختلافات الكلية للمتغير التابع نحصل على:

$$\frac{\sum(Y-\bar{Y})^2}{\sum(Y-\bar{Y})^2} = \frac{\sum(Y'-\bar{Y})^2}{\sum(Y-\bar{Y})^2} + \frac{\sum(Y-Y')^2}{\sum(Y-\bar{Y})^2}$$

علماً أن نسبة الاختلافات المفسرة إلى الاختلافات الكلية تعرف بمعامل التحديد R^2 أي أن:

$$R^2 = \frac{\sum(Y'-\bar{Y})^2}{\sum(Y-\bar{Y})^2} = 1 - \frac{\sum(Y-Y')^2}{\sum(Y-\bar{Y})^2}$$

(6)

يتضح من المعادلة (6) أعلاه أن أقصى قيمة (Maximum value – Valeur maximale) والتي يمكن لمعامل التحديد أن يأخذها هي الواحد الصحيح $R^2 = 1.0$، وذلك عندما $Y = Y'$ حيث تنعدم الأخطاء في التوقع (التنبؤ)، كما وأن أدنى قيمة (Minimum value – Valeur minimale) لمعامل التحديد هي الصفر $R^2 = 0.0$ عندما $Y' = \bar{Y}$ حيث يكون $\bar{Y}$ أفضل قيمة يمكننا التنبؤ بها في حالة انعدام العلاقة بين المتغيرينX و Y .

تجدر الإشارة أخيراً إلى أن الجذر التربيعي لمعامل التحديد يعطي معامل الارتباط $\sqrt{R^2} = \pm r$، كما وأن الإشارة (Sign -Signe) الجبرية لمعامل الارتباط تكون مساوية للإشارة الجبرية لمعامل الانحدار لمعرفة اتجاه (Direction - Direction) العلاقة بين المتغيرين، وبالحصول على معامل التحديد لمعرفة نسبة الاختلاف المفسرة من الاختلافات الكلية.

معامل الارتباط البسيط (The simple correlation coefficient – Coefficient de correlation simple):

يقصد بالارتباط بشكل عام وجود علاقة، لكن المعنى الإحصائي للارتباط (Correlation - Correlation)، هو أوسع من كلمة علاقة (Relationship - Relation)، لأنه يعني دراسة التغاير (Covariance - Covariance) بين المتغيرات، أي دراسة العلاقة بين تراتيب قيم المتغير X وتراتيب قيم المتغير Y.

دعنا نفترض للتبسيط أننا نتعامل مع المتغيرات في شكل انحرافات عن الوسط الحسبي (Deviation scores - Deviations de la moyenne) حيث تكون $b_0 = 0.0$، وبالتالي يمكننا كتابة معادلة الانحدار في شكل انحرافات كالآتي:

$$y' = bx$$

ونظراً لأن معامل التحديد يساوي:

$$R^2 = 1 - \frac{\sum(Y-Y')^2}{\sum(Y-\bar{Y})^2}$$

إذن:

$$R^2 = 1 - \frac{\sum (y - bx)^2}{\sum y^2}$$

$$R^2 = \frac{\sum y^2 - \sum (y - bx)^2}{\sum y^2}$$

$$R^2 = \frac{\sum y^2 - \sum y^2 + 2b \sum xy - b^2 \sum x^2}{\sum y^2}$$

ونظراً لأنّ:

$$b = \frac{\sum xy}{\sum x^2}$$

$$\sum xy = b \sum x^2$$

$$b \sum xy = b^2 \sum x^2$$

إذن:[12]

$$R^2 = b^2 \frac{\sum x^2}{\sum y^2} = b . \frac{\sum xy}{\sum x^2} . \frac{\sum x^2}{\sum y^2} = b \frac{\sum xy}{\sum x^2}$$

وباستبدال b بقيمتها نحصل على:

$$R^2 = \frac{\sum xy}{\sum x^2} . \frac{\sum xy}{\sum y^2}$$

$$R^2 = b_{yx} . b_{xy}$$ [13]

$$r = \sqrt{b_{yx} . b_{xy}} = \frac{\sum xy}{\sqrt{\sum x^2 . \sum y^2}}$$

وتجدر الإشارة أنه يوجد إضافة إلى الصيغة الموضحة أعلاه، صيغ أخرى لإيجاد معامل الارتباط أهمها $r = \frac{\sum Z_X Z_Y}{n-1}$ ، والتي تستعمل في حالة التعامل مع القياسات في شكل وحدات معيارية (Standard score – Scores standards).

---

[12] تشير المعادلة $R^2 = b\frac{\Sigma xy}{\Sigma y^2}$ إلى أن زيادة أي متغير مستقل إلى معادلة الانحدار سيعني بالضرورة ارتفاع قيمة معامل التحديد، فلو وُجد متغيرين، في معادلة الانحدار مثلاً، لكان معامل التحديد $R^2 = \frac{b_1 \Sigma x_1 y + b_2 \Sigma x_2 y}{\Sigma y^2}$ انظر ص 115-116 من هذا المؤلف.

[13] تشير b_{yx} إلى معامل انحدار المتغير التابع Y على المتغير المستقل X في معادلة الانحدار $Y = F(X)$ في حين تشير b_{xy} إلى معامل المتغير التابع X على المتغير المستقل Y في معادلة الانحدار $X = b_0 + b_1 Y$. حيث يُعامل كلٍ من المتغيرين على أنه مستقل تارةً وتابع تارة أخرى

تجدر الإشارة أخيراً إلى ضرورة اخذ الحيطة والحذر، عند تفسير معامل الارتباط البسيط بين المتغيرات الاقتصادية. فالارتباط لا يعني السببية (Causation - Causalité)، أي أن وجود ارتباط بين متغيرين لا يعني وجود سبب (Cause - Cause) ونتيجة (Effect - Effet) فكثيراً ما نحصل على ارتباط زائف (Spurious correlation – Correlation erronée) بين متغيرين تنعدم قيمته أو تزيد بإدخال الرقابة الإحصائية (Statistical control – Contrôle statistique) فلو فرضنا على سبيل المثال إن معامل الارتباط بين رواتب المدرسين وبين استهلاك النبيذ خلال حقبة زمنية معينة كان $r = 0.98$ فهذا لا يعني أنّ المدرسين يشربون النبيذ، كما وأن وجود الارتباط لا يعني أن استهلاك النبيذ يتزايد مع تزايد رواتب المدرسين فالعلاقة الموجودة هي علاقة زائفة ناتجة عن تأثر كلا المتغيرين بمتغير ثالث وهو الزمن و بشكل أدق فإن كلا المتغيرين يتأثران بازدياد الدخل القومي، وعندما نستبعد اثر المتغير الخارجي (الزمن) فإن العلاقة بين المتغيرين قد تنعدم و تصل إلى الصفر[14].

فترة الثقة (Confidence interval – Intervalle de confiance):

يلجأ الباحث عادة إلى استعمال بيانات إحصائية من العينة وذلك ليتمكن من عمل استدلالات عن معالم المجتمع الإحصائي. علماً أنه بناء على نظرية النزعة المركزية (The central limit theorem – Le théorème de la limite centrale) فإن توزيع المعاينة (The sampling distribution - Echantillonnage) لأكثر الظواهر يكون معتدلاً (Normally distributed – Normalement distribuée).

لنفترض أنه لدينا عينة عشوائية مكوّنة من (n) المشاهدات (Observations - Observations)، ذات وسط حسابي $\bar{x}$ وانحراف معياري s_X، فهل يتساوى الوسط الحسابي لهذه العينة، مع الوسط الحسابي μ للمجتمع الإحصائي؟ نلاحظ أن قيمة $\bar{x}$ نادراً ما تساوي قيمة μ، بسبب خطأ المعاينة، إلّا أننا نتوقع أن تكون قيمة $\bar{x}$ قريبة جداً من قيمة μ. ولو أخذنا عينات عشوائية معادة ذات حجم (n) (Repeated random samples size n – Echantillons aléatoires répétés de taille n) من نفس المجتمع الإحصائي ثم أوجدنا الوسط الحسابي لكل عينة، فسنجد أن الوسط الحسابي للأوساط الحسابية للعينات يساوي الوسط الحسابي للمجتمع الإحصائي (لذلك يقال بأن الوسط الحسابي للعينة هو تقدير غير متحيز لقيمة المعلم في المجتمع الإحصائي)، كما وأن توزيع الأوساط الحسابية للعينات تكون موزعة توزيعاً معتدلاً حول μ، وبانحراف معياري $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$، بحيث أن 95% من الأوساط الحسابية للعينات ستقع ضمن المدى μ C $1.96\ \sigma_{\bar{X}}$ لذلك فإذا علمنا قيمة $\bar{X}$ من العينة فبإمكاننا القول أن μ للمجتمع الإحصائي ستقع ضمن $\bar{X}$ C $1.96\ \sigma_{\bar{X}}$، علماً أن القيمة 1.96 هي القيمة الحرجة(critical value – Valeur critique) المعروفة بالقيمة الجدولية (table value – Valeur tabulée)، ويتم الحصول عليها من جداول التوزيع المعتدل (Areas under the standard normal curve – Surface sous la courbe normale). الجدير بالذكر أنه إذا قررنا أن μ تقع ضمن المدى $\bar{X}$ C $1.96\ \sigma_{\bar{X}}$ فحينئذ نكون واثقين بنسبة 95% من قرارنا الإحصائي، ونكون مخطئين بنسبة 5% في قرارنا الإحصائي. وبالمثل فإن فترة الثقة $\bar{X}$ C $2.58\ \sigma_{\bar{X}}$ تمثل فترة الثقة بنسبة 99%، علماً أنه باستطاعة الباحث توسيع أو تضييق فترة الثقة وذلك بتغيير حجم العينة أو/ و تغيير مستوى المعنوية(the level of significance – Le niveau de signification)[15].

---

[14] انظر معامل الارتباط الجزئي والنصف جزئي ص :121-123

[15] يعرف احتمال ارتكاب الباحث للخطأ من النوع الأول (Type I error – Erreur de type 1) بمستوى المعنوية ألفا α.
ويتخذ الباحث عادة المستوى 1% أو 5% والذي يمثل الحد الأقصى الذي يمكن للباحث أن يتحمل فيه ارتكاب الخطأ من النوع الأول. فلو استعمل الباحث المستوى 5% فإنه بذلك يكون واثقاً بنسبة 95% من أنه سيتخذ القرار الإحصائي السليم في رفض فرض العدم. والجدير

آخذين في الاعتبار أنه إذا كان حجم العينة صغيراً (أقل من 30 مشاهدة) فحينئذ يتوجب على الباحث استبدال القيم الحرجة 1.96 و 2.58 بقيم أخرى مناسبة يتم الحصول عليها من جدول توزيع استودينت(student's t-distribution – Distribution de Student t)، باستخدام مستوى معنوية ودرجات حرية (degree of freedom – Degré de liberté) مناسبة[16].

يهتم الباحث في مجال الاقتصاد القياسي بشكل خاص بإيجاد فترات الثقة لمعاملات الانحدار b_i، وللقيمة المتوقعة في المجتمع الإحصائي $E(Y/X)$ ولا شك أنه حتى يتمكن من صياغة فترة الثقة لمعامل الانحدار، لا بد له من معرفة الخطأ المعياري لتوزيع المعاينة لمعامل الانحدار (The standard error of the sampling distribution of b – L'écart-type de la distribution d'échantillonnage)، علماً أن الخطأ المعياري، لا يعدو عن كونه الانحراف المعياري لتوزيع المعاينة. ولإيجاد تباين توزيع المعاينة لمعامل الانحدار b، نلاحظ أن بالإمكان تعريف تباين معامل الانحدار على أنه:

$$E\left[b - E(b)\right]^2 = \frac{\Sigma(b-\beta)^2}{M}$$

حيث تمثل M عدد العينات. ويلاحظ في المعادلة أعلاه أن كلاً من β و M غير معلومة (unknown - Inconnue) للباحث، كما ويفترض أن تكون M كبيرة جداً. لذلك يمكن القول أن المعادلة أعلاه صعبة الإيجاد من الناحية العملية ويستعاض عنها بالمعادلة العملية:

$$\sigma_b^2 = \frac{SS_{\mathrm{Re}\,sd}}{n-k-1} \cdot \frac{1}{\sum x^2}$$

حيث تمثل n عدد المشاهدات في العينة، وترمز K إلى عدد المتغيرات المستقلة في معادلة الانحدار[17]. وبما أننا حصلنا على تباين توزيع المعينة لمعامل الانحدار، فقد أصبح بالإمكان اختبار

---

[16] يستخدم توزيع استودينت (student's t – distribution – distribution de student t) على متغير عشوائي يتوزع توزيعاً معتدلاً، ويكون الوسط الحسابي في المجتمع غير معلوم (أو يفترض أنه معلوم)، كما وأن التباين يقدر من العينة. ويتميز توزيع اسودينت –T على التوزيع المعتدل (Normal distribution – Distribution normale) في أن الأخير هو توزيع فريد (single distribution – distribution simple)، بينما نجد أن توزيع استودينت – T هو عائلة من التوزيعات (Family of distributions – famille de distributions)، يعتمد كل منها على درجات الحرية المناسبة. أما درجات الحرية (degree of freedom – degrés de liberté) فيقصد بها عدد الحدود (terms - termes) التي يمكن أن تتحرك بحرية في مجموعة من البيانات. فلو فرضنا أن مجموع خمسة أرقام هو 24, وأن أربعة أرقام من أصل الخمسة ممكن أن تكون أي شئ مثل 6 , 4 , 2 و 10 ومجموعها 22, حينئذٍ يجب أن تكون قيمة الرقم الخامس 2 وذلك حتى يصبح مجموع الأرقام الخمسة 24 , وبالتالي يمكننا القول أن أحد الأرقام ليس مستقلاً عن قيم الآخرين بينما الآخرين مستقلين، فنقول في هذا أننا وضعنا قيداً واحد (one restriction – une seule restriction) على البيانات، فنفقد لذلك درجة حرية واحدة (one degree of freedom – un seul degré de liberté) وتكون درجات الحرية لمثالنا السابق $dF = 5 - 1 = 4$. وهذا ما يفسر استخدام $n-1$ بدلاً من n في إيجاد التباين للعينة، حيث أن مجموع الانحرافات عن الوسط الحسابي يساوي صفراً، فنفقد درجة واحدة ويكون التباين:

$$S_X^2 = \frac{\Sigma\left(X - \overline{X}\right)^2}{n-1}$$

[17] لاحظ أنّ:

$$b = \sum \omega Y = \sum \omega(\alpha + \beta X + u) = \alpha \sum \omega + \beta \sum \omega X + \sum \omega U$$

$$\beta = b \subset Z_c \sigma_b \qquad (1)$$

$$\beta = b \subset t_c \sigma_b \qquad (2)$$

حيث تستعمل الصيغة (1) إذا كان حجم العينة (n) كبيراً (أكثر من 30 مشاهدة)، بينما تستعمل الصيغة (2) في حالة العينات الصغيرة (أقل من 30 مشاهدة). كذلك تجدر الإشارة إلى أنه باستطاعة الباحث أيضاً تحديد فترة الثقة للقيمة المتوقعة $E(Y/X)$ ، وتحديد فترة الثقة للقيمة الفعلية Y كالآتي:

$$E(Y/X) = \mu_{XY} = \alpha + \beta X = \overline{Y} + b(X - \overline{X})$$

$$\sigma^2_{\mu_{xy}} = \sigma^2\left[\overline{Y} + b(X - \overline{X})\right]$$

$$\sigma^2_{\mu_{xy}} = \sigma^2(\overline{Y}) + (X - \overline{X})^2 \sigma^2(b)$$

$$\sigma^2_{\mu_{xy}} = \frac{\sigma^2_{yx}}{n} + (X - \overline{X})^2 \frac{\sigma^2_{yx}}{\sum X^2}$$

$$\sigma_{\mu_{xy}} = \sigma_{yx}\sqrt{\frac{1}{n} + \frac{(X - \overline{X})2}{\sum X^2}} = \sigma_{yx}\sqrt{\frac{1}{n} + \frac{X^2}{\sum X^2}}$$

أما فترة الثقة فيمكن صياغتها كالآتي:

$$\grave{Y} \subset Z_c \sigma_{\mu_{yx}}$$

اختبار الفروض (Testing hypotheses – Test d’hypothèses):

يعتمد الاقتصاد القياسي على البحث العلمي (Scientific research – Recherche scientifique)، ولا شك بأن الخطوة الأولى في البحث العلمي، هي أن يحدد الباحث مشكلة البحث (Problem- Idea - Problèmatique). والجدير بالذكر، أنه بعد أن يعرف الباحث ماذا يريد أن يبحث فعلاً، عليه حينئذ أن يصيغ فروض بحثه (Formulating Testable hypothesis – Formulation des hypothèses testables)، فالباحث يلاحظ ظاهرة (Phenomenon - Phénomène) ما، ومن ثم يتوقع مسببات ونتائج لهذه الظاهرة. لذلك يقال أن فروض البحث (Research hypothesis – Hypothèse de recherche)، هي في الحقيقة توقعات ورهان (Prediction and betting – Prévision et paris) وعلى الباحث أن يلتزم بقواعد البحث العلمي. فصياغة فروض البحث هي من قواعد البحث العلمي التي تهدف إلى تقليل الأخطاء (Errors

---

$$\sum \omega = \frac{\sum X}{\sum X^2} = 0$$

$$\sum \omega X = \sum \frac{X}{\sum X^2} . X = 1$$

إذن:

$$b = \beta + \sum \omega U$$

$$b - \beta = \sum \omega U$$

$$E\left[b - \beta\right]^2 = \sum \omega^2 E(U)^2 = \sum\left(\frac{X}{\sum X^2}\right)^2 . \sigma^2_{xy} = \frac{1}{\sum X^2} \cdot \frac{SS_{\mathrm{Re\ sd}}}{n - k - 1}$$

تجدر الإشارة إلى أن الباحث في بداية الأمر يفترض فرضاً عاماً وواسعاً (General - Générale) وإذا كان هذا الفرض جيداً، فباستطاعة الباحث حينئذٍ صياغة فروض البحث والتي يجب أن تتوافر فيها المواصفات التالية :

أ- يجب أن تصاغ فروض البحث في شكل جمل استفهامية.
ب- يجب أن تربط فروض البحث بن متغيرين أو أكثر.
ت- يجب أن تتضمن فروض البحث على مفهوم ضمني مؤداه إمكانية قياس المتغيرات الاقتصادية وبالتالي إمكانية إجراء اختبارات إحصائية على العلاقات قيد البحث.
ث- يجب أن تكون فروض البحث محددة، فعندما تكون التوقعات محددة يصغر احتمال الحصول على نتائج أو فروقات جوهرية في البحث عن طريق الصدفة[18].

الجدير بالذكر، أنه يتوجب على الباحث التمييز بين فروض البحث والفروض الإحصائية (Statistical hypotheses – Hypothèse statistiques). ففروض البحث تكون عامة، ومبنية على نظرية علمية، أو مبنية على نتائج بحوث سابقة، أو مبنية على أسس منطقية. ومثل هذه الفروض كما ذكرنا، تتضمن توقعات لنتائج البحث، ويستدل الباحث منها إلى فروض إحصائية، قابلة للاختبار الإحصائي (Testable hypotheses – Hypothèses testables) .

تصاغ الفروض الإحصائية لتقييم فروض البحث، علماً أن الفروض الإحصائية، هي تعبير عن واحد أو أكثر من معالم المجتمع الإحصائي، التي سحبت منها العينة . وفرض العدم (The null hypothesis: H_0 – Hypothèse nulle: H_0) والفرض البديل (The alternative hypothesis: H_1 – Hypothèse alternative: H_1)، هما شكلان من الفروض الإحصائية. فعلى سبيل المثال، تقترح النظرية الاقتصادية لعرض سلعةٍ ما، وجود علاقة إيجابية بين العرض والثمن، ونظراً لأن النظرية لم تحدد فيما إذا كانت العلاقة خطية (Linear - Linéaire) أو غير الخطية (Nonlinear – Non linéaire)، فإننا سنفترض أن العلاقة خطية وبالتالي يمكننا صياغة معادلة الانحدار كالآتي[19]:

$$\hat{Y} = b_0 + b_1 X$$

وتتلخص مشكلة الباحث، في تقدير العلاقة بين العرض والثمن، وفي تقدير قيم المعالم α و β، من خلال معرفتنا بمعاملات الانحدار b_0 و b_1 التي حصلنا عليها من بيانات العينة. وهنا نلاحظ أن الباحث، وقبل جمع البيانات، يتوقع أن تكون قيمة b_0 موجبة أو صفراً، حيث تعني القيمة الموجبة للثابت b_0 (Intercept - Constante) أنه توجد كمية معروضة من السلعة في السوق حتى ولو كان السعر صفراً $X = 0.0$، بمعنى أن الباحث لا يتوقع أن تكون b_0 سالبة. ولو حدث أن حصل الباحث على قيمة سالبة، للثابت b_0، فعليه إهمالها لأنها لا تعني شيئاً بالنسبة له. آخذين في الاعتبار، أنه إذا كانت مشكلة البحث تتضمن لتحديد مرونة (Elasticity - Elasticité) العرض، فحينئذ تلعب الإشارة السلبية لمعامل الانحدار b_0 دوراً هاماً في تحديد المرونة، لأنها عندما تدخل في احتساب

---

[18] Fred N. kerlinger., " Foundation of behavioral Research"- 2nd ed. holt-Rinehatr and Winston, inc., 1973,pp: 16-22

[19] يقصد بالخطية أعلاه، أن المعادلة خطية في المعالم α و β. بمعنى أن هذه المعالم مرفوعة إلى القوة الأولى (α و β are raised to the first power).

0

$$\frac{\frac{\Delta Y}{Y}}{\frac{\Delta X}{X}} = \frac{\text{المرونة= التغير النسبي في الكمية}}{\text{التغير النسبي في الثمن}}$$

$$\eta = \frac{\Delta Y}{Y} \cdot \frac{X}{\Delta X}$$

$$\eta = \frac{\Delta Y}{\Delta X} \cdot \frac{X}{Y}$$

وعلى افتراض أن $\overline{X}$ تمثل السعر X في حين $\overline{Y}$ تمثل الكمية المعروضة Y ، إذن[21]:

$$\eta = b_1 \cdot \frac{\overline{X}}{\overline{Y}}$$

$$\eta = b_1 \cdot \frac{\overline{X}}{b_0 + b_1\overline{X}}$$

$$\eta = \frac{b_1\overline{X}}{b_0 + b_1\overline{X}}$$

وبناء على الصيغة أعلاه يمكننا القول أنه:

- إذا كانت $b_0 = 0.0$ فحينئذٍ المرونة تساوي الواحد الصحيح، (عرض متكافئ المرونة) حيث تتساوى التغيرات النسبية للعرض والسعر، ويكون العرض بذلك متكافئ المرونة حيث يؤدي تغيير الثمن إلى تغير الكمية المعروضة بنفس النسبة، فلو ارتفع الثمن إلى الضعف فإن الكمية المعروضة ترتفع إلى الضعف.

- إذا كانت b_0 سالبة فحينئذٍ المرونة أكبر من الواحد الصحيح حيث يزيد التغير النسبي في العرض عند التغير النسبي في السعرفيكون العرض بذلك مرناً (Elastic - Elastique). ومن الواضح أن المرونة تزداد كلما كانت النتيجة أكثر بعداً عن الواحد الصحيح، بحيث أن تغير طفيف في الثمن يحدث تغيراً كبيراً في الكمية المعروضة.

- إذا كانت b_0 موجبة، فحينئذٍ تكون المرونة أقل من الواحد الصحيح، حيث يقل التغير النسبي في العرض عن التغير النسبي في السعرفيكون العرض بذلك قليل المرونة (inelastic - inélastique). ومن الواضح أن المرونة تزداد ضعفاً كلما كانت النتيجة أقل بكثير من الواحد الصحيح، حيث لا تتأثر الكمية المعروضة كثيراً بتغيرات الثمن.

[20] يقصد بمرونة العرض درجة استجابة العرض للتغير في العامل الذي يؤثر عليه. للتوسع في ذلك انظر: د. إسماعيل محمد هاشم ص: 237-236

[21] نظراً لأن المرونة تتغير عند كل نقطة في الدالة لذلك تأخذ الأوساط الحسابية $\overline{X}$ و $\overline{Y}$ بدلاً من X و Y

أما بالنسبة لميل خط الانحدار b_1 فيتوقع الباحث أن يكون الميل موجباً، لأن خط انحدار العرض على الثمن يكون صاعداً نحو الأعلى، حيث تفترض النظرية الاقتصادية وجود علاقة طردية بين العرض والثمن.

لا شك أنه بعد الحصول على التقديرات b_0 و b_1 يتوجب على الباحث اختبار معنوية (Significance - signification) التقديرات، حيث يصيغ الباحث فرضية العدم، ومؤداها أن البيانات الإحصائية، عن الكمية والسعر، هي عبارة عن عينة عشوائية، مسحوبة من مجتمع إحصائي لا يوجد فيه انحدار للكمية المعروضة على السعر:

$$H_0 : \beta = 0$$

علماً أن فرض العدم غالباً ما يكون في الاتجاه المعاكس لفرض البحث، لذلك يهدف الباحث عادة إلى رفض فرض العدم، أو إبطاله، بالاختبار الإحصائي، وأخذ الفرض البديل له:

$$H_1 : \beta \neq 0$$

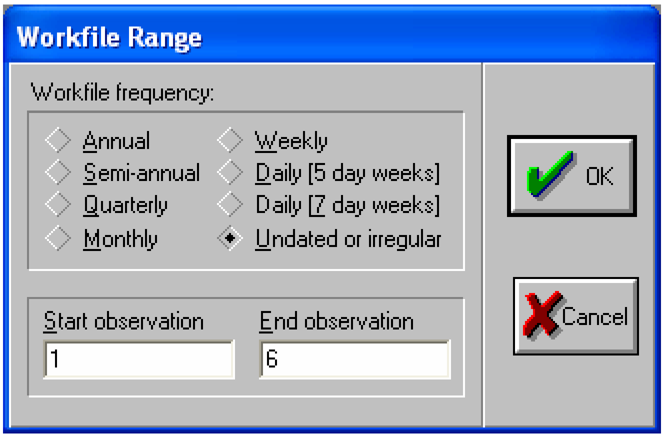
والذي ينص على وجود علاقة بين الكمية المعروضة والسعر في المجتمع الإحصائي. علماً أنه باستطاعة الباحث اختبار معامل التحديد بدلاً من معامل الانحدار[22]. وتجدر الإشارة أخيراً إلى أن فرض العدم هو تعبير يتضمن واحد أو أكثر من المقاييس الخاضعة لاختبار إحصائية وهو بالتالي الفرض الذي يمكنه رفضه لكن لا يمكن برهنته.

مثال انتقال أثرالتضخم العالمي الى الأقتصاد الكويتي :(اللجنة الأقصادية لغرب اسيا 1970-1978 ص 136)

استخدام برنامج EViews في تحليل مثال عن انتقال أثر التضخم العالمي الى الأقتصاد الكويتي :

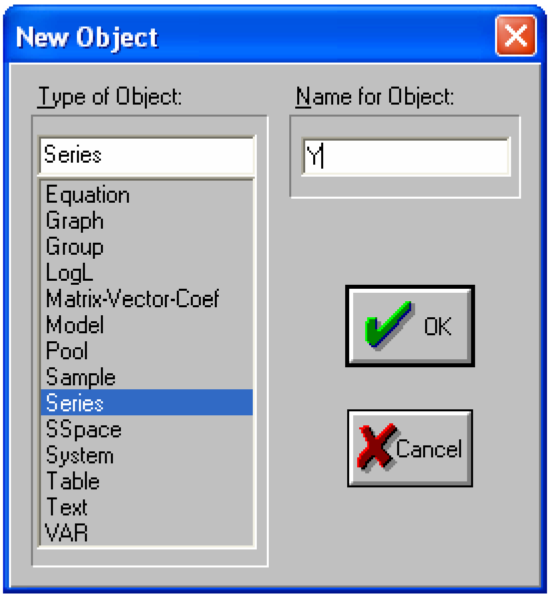
<u>أولا</u>: نختار File \ New \ Workfile

<u>ثانيا</u>:

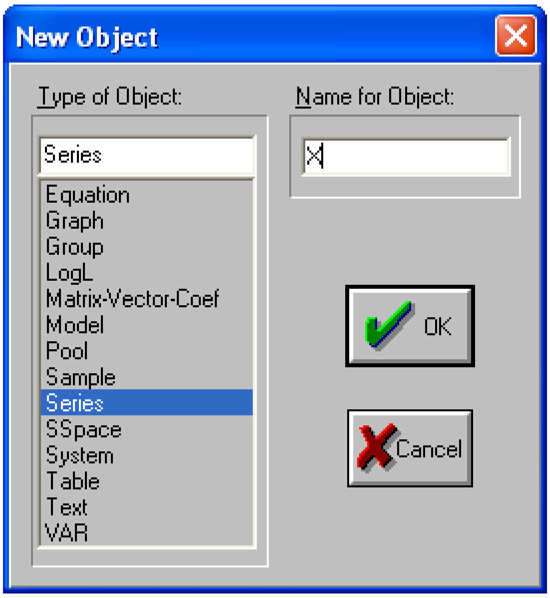


ثالثا: نعرف المتغير التابع File \ Object \ New Object

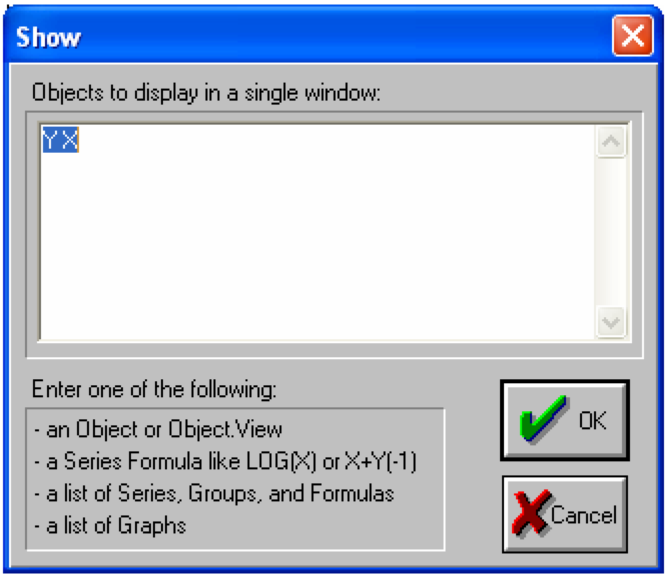
[22] تجدر الإشارة إلى أن استخدام إحصائية F (F- Test) في اختبار معامل التحديد هي أفضل من اختبار معامل الانحدار وخاصة في حالة الانحدار المتعدد حيث يًخشى من وجود النماذج المتكافئة (Equivalent Models) حيث يكون إحدى المتغيرات المستقلة دالة في متغير مستقل آخر (Linear dependency) . علماً أن توزيع F (F- distributions) هو التوزيع النظري لنسبة التباين بين مجتمعين وقد وًجد على يد السير فيشر (Fisher) في أوائل العشرينات من القرن الحالي لكن طُور فيما بعد بشكل يُسهل استعماله وسُمي على شرف فيشر.

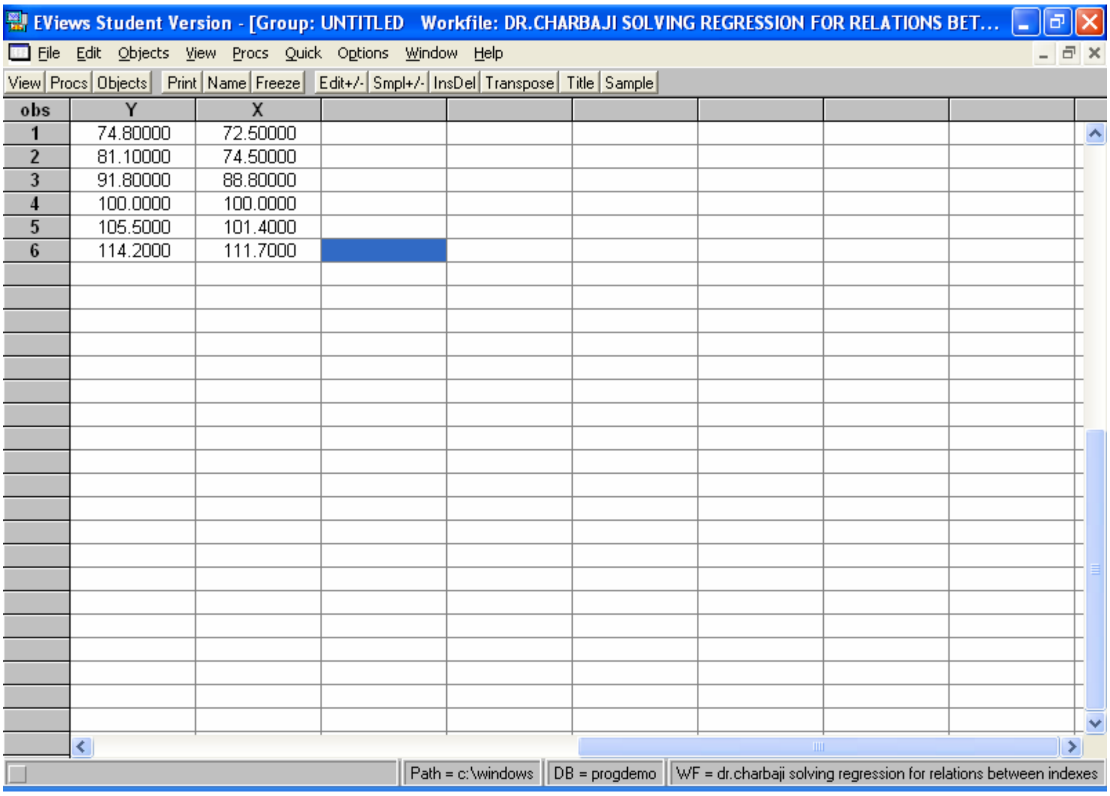


خامسا: نعرف المتغيرالمستقل:

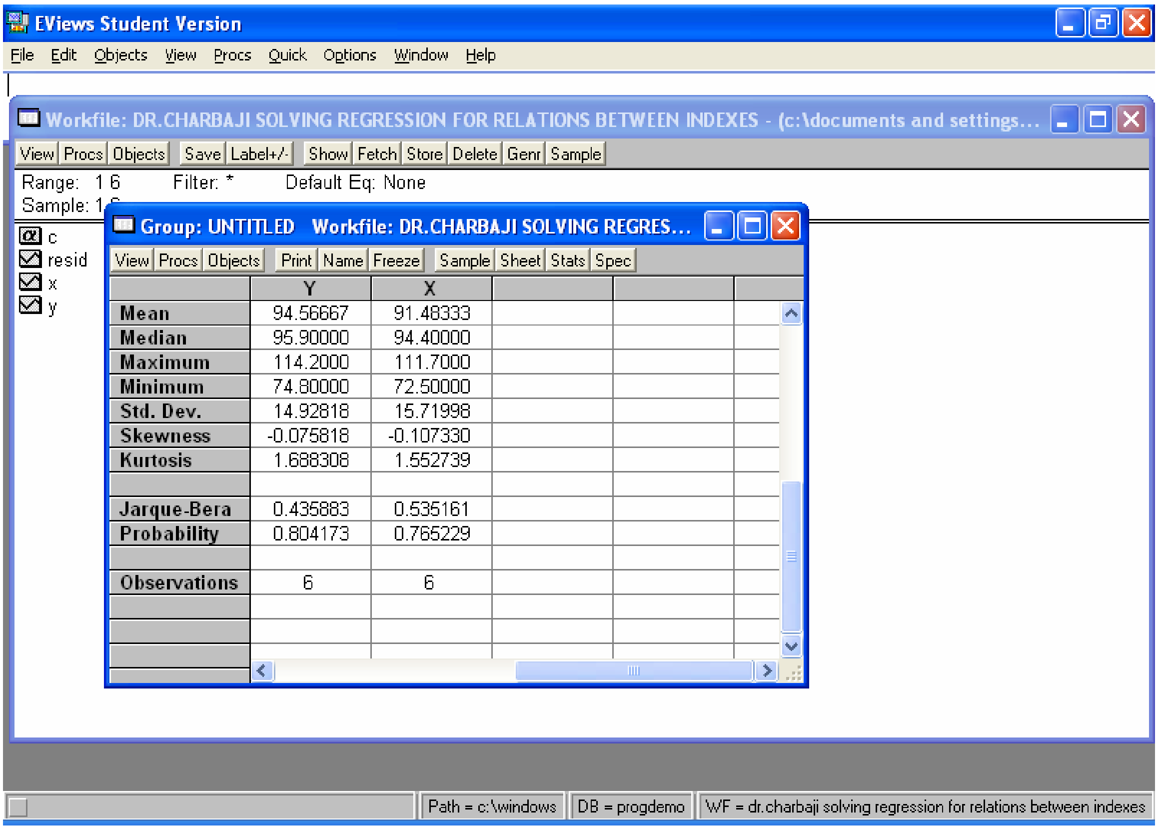


سادسا: نختار المتغيرين معا: **Control \ Show \ Edit** +/-


Show
Objects to display in a single window:
Y X
Enter one of the following:
- an Object or Object.View
- a Series Formula like LOG(X) or X+Y(-1)
- a list of Series, Groups, and Formulas
- a list of Graphs
OK
Cancel

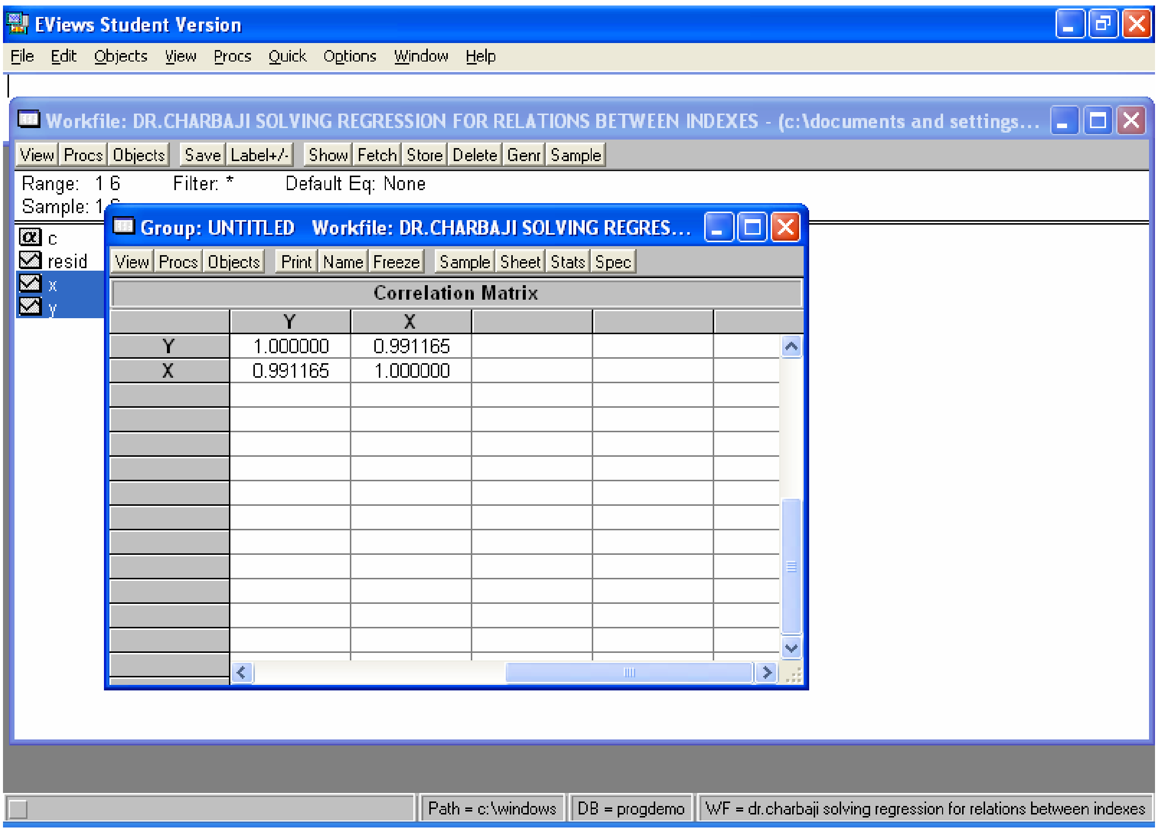

EViews Student Version - [Group: UNTITLED Workfile: DR.CHARBAJI SOLVING REGRESSION FOR RELATIONS BET...
File Edit Objects View Procs Quick Options Window Help
View Procs Objects Print Name Freeze Edit+/- Smpl+/- InsDel Transpose Title Sample
obs Y X
1 74.80000 72.50000
2 81.10000 74.50000
3 91.80000 88.80000
4 100.0000 100.0000
5 105.5000 101.4000
6 114.2000 111.7000
Path = c:\windows
DB = progdemo
WF = dr.charbaji solving regression for relations between indexes

View \ Descriptive Statistics \ Individual Samples



	Y	X
Mean	94.56667	91.48333
Median	95.90000	94.40000
Maximum	114.2000	111.7000
Minimum	74.80000	72.50000
Std. Dev.	14.92818	15.71998
Skewness	-0.075818	-0.107330
Kurtosis	1.688308	1.552739
Jarque-Bera	0.435883	0.535161
Probability	0.804173	0.765229
Observations	6	6

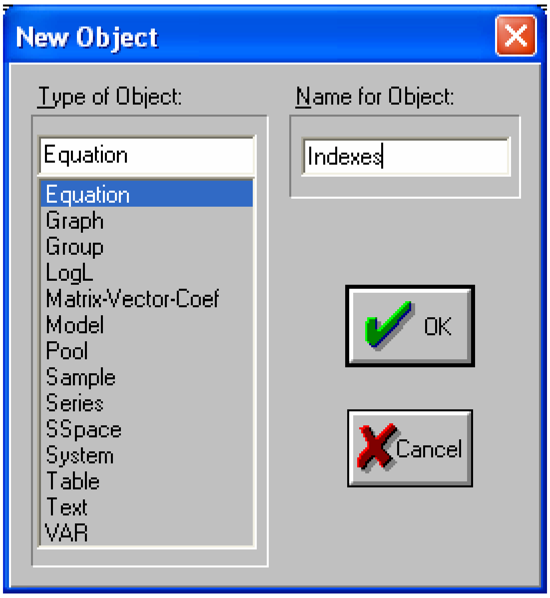
View \ Descriptive Statistics \ Correlation Matrix

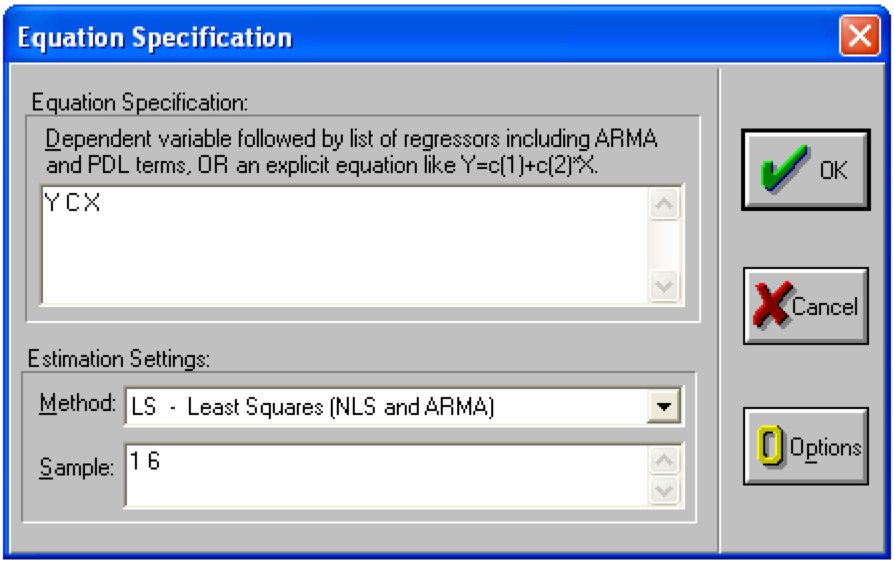


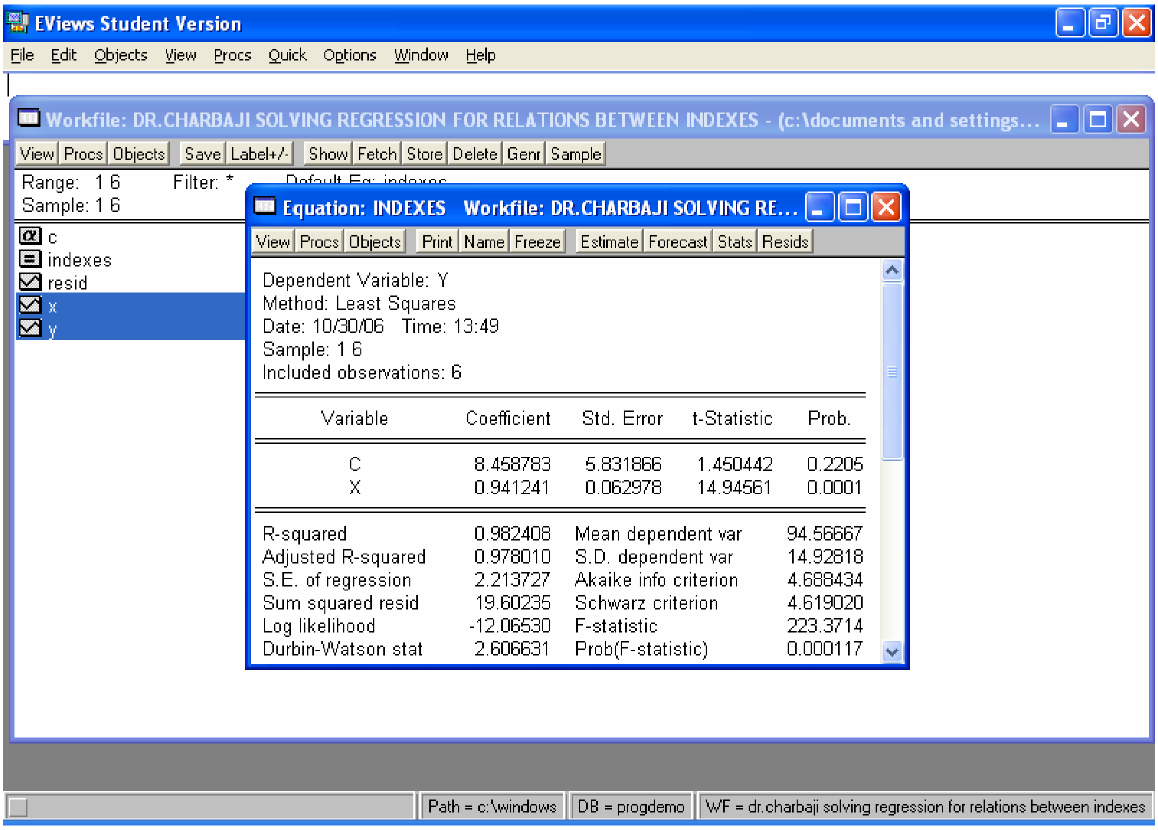
Correlation Matrix

	Y	X
Y	1.000000	0.991165
X	0.991165	1.000000

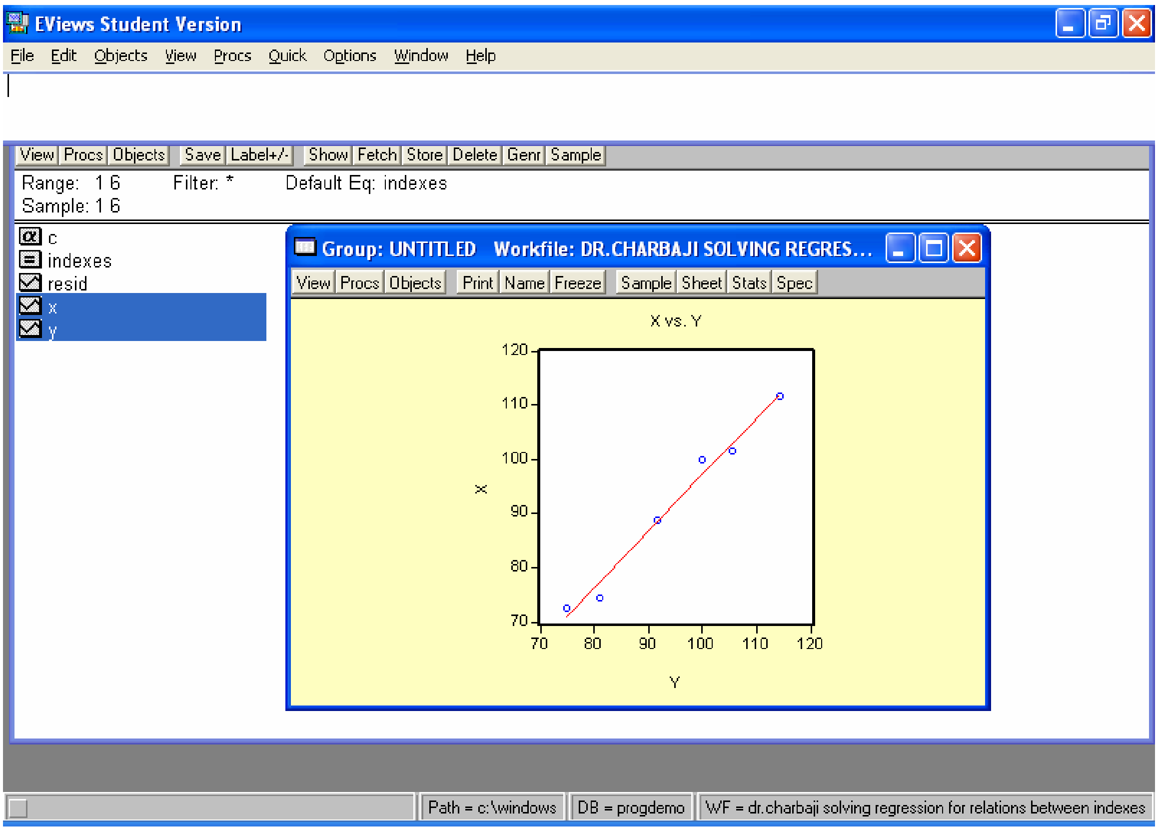
Object \ New Object \ Equation



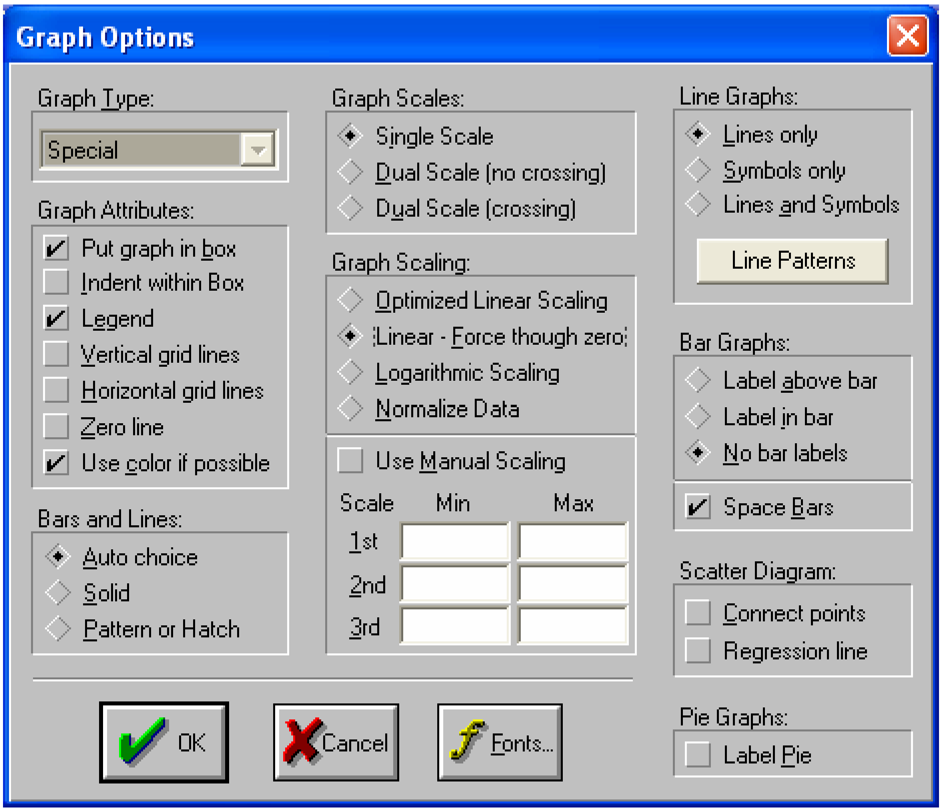


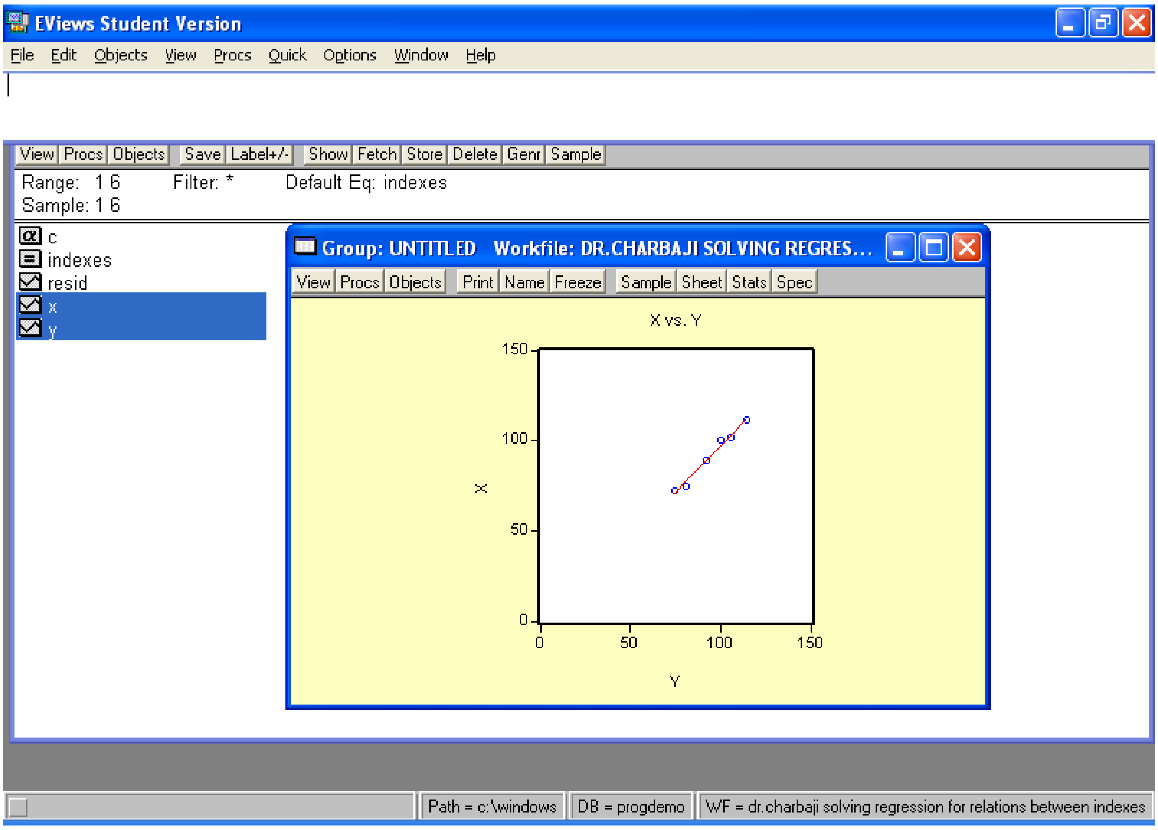


View \ Graph \ Scatter \ Scatter With Regression

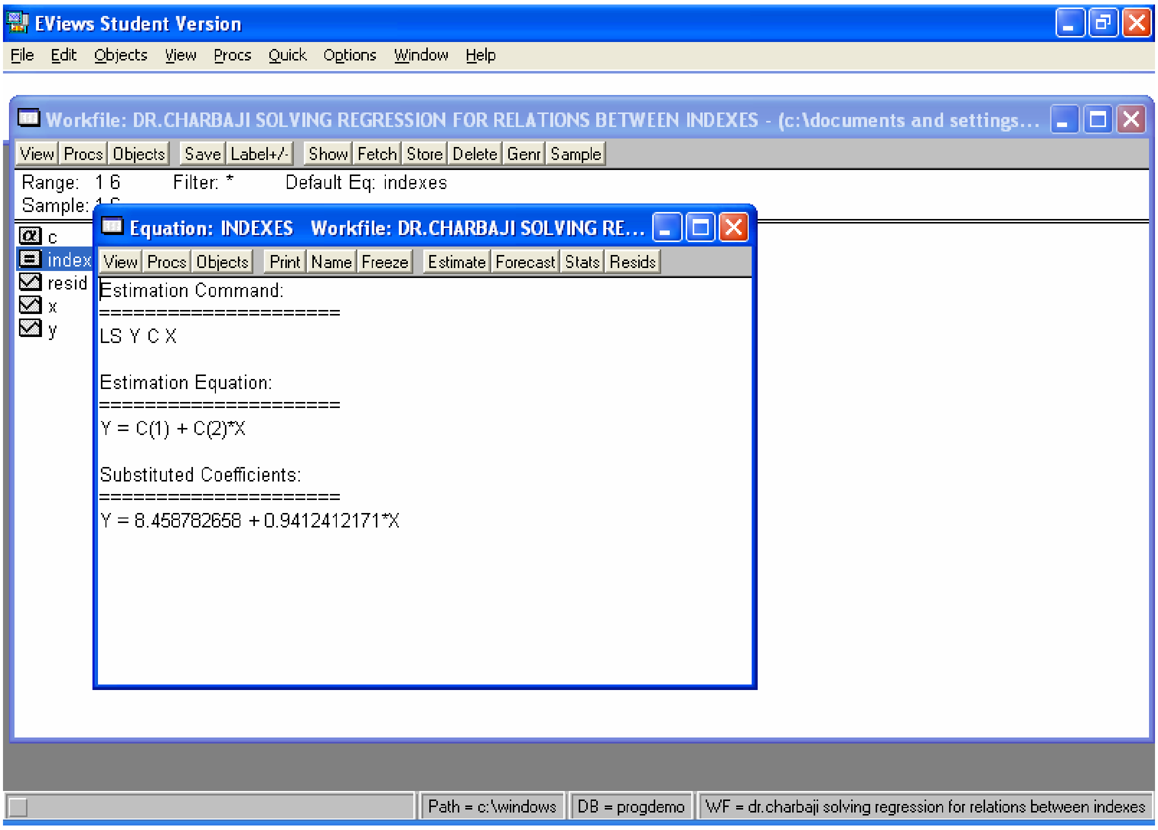


Click the right mouse button on graph \ Choose Options \ Linear Force Thru Zero

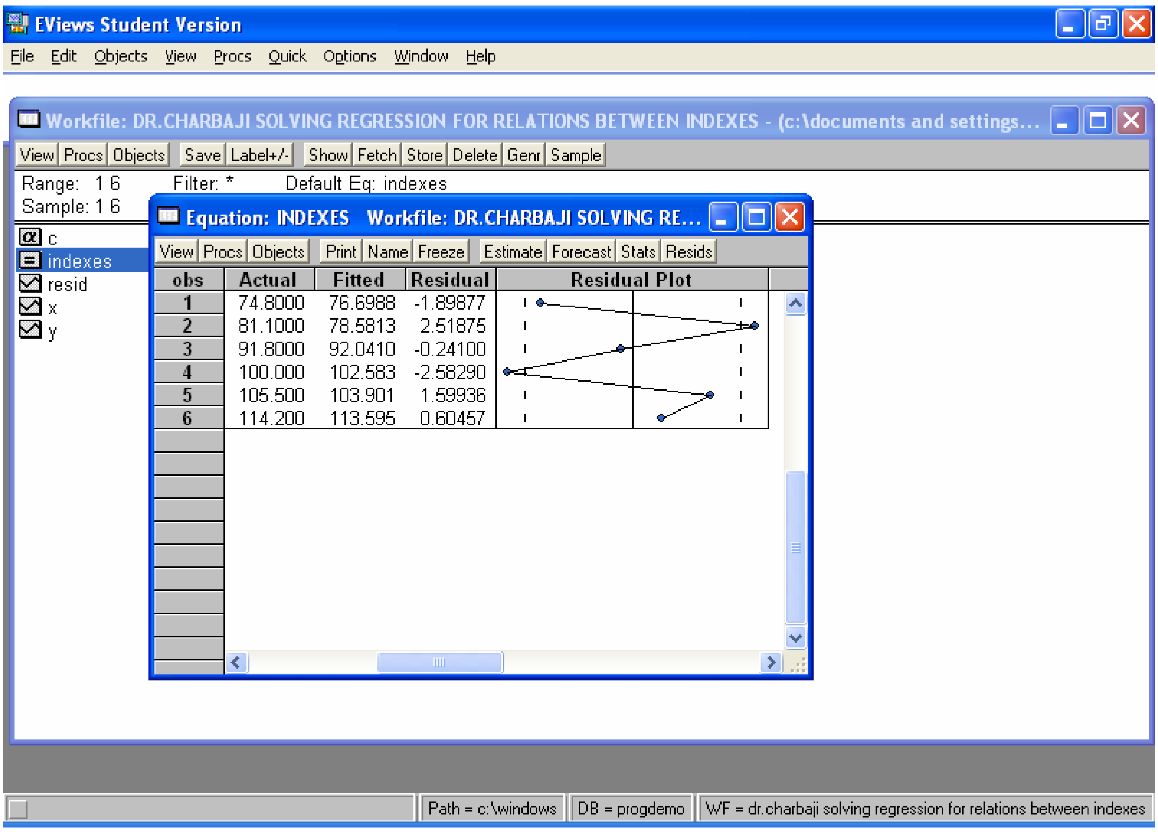




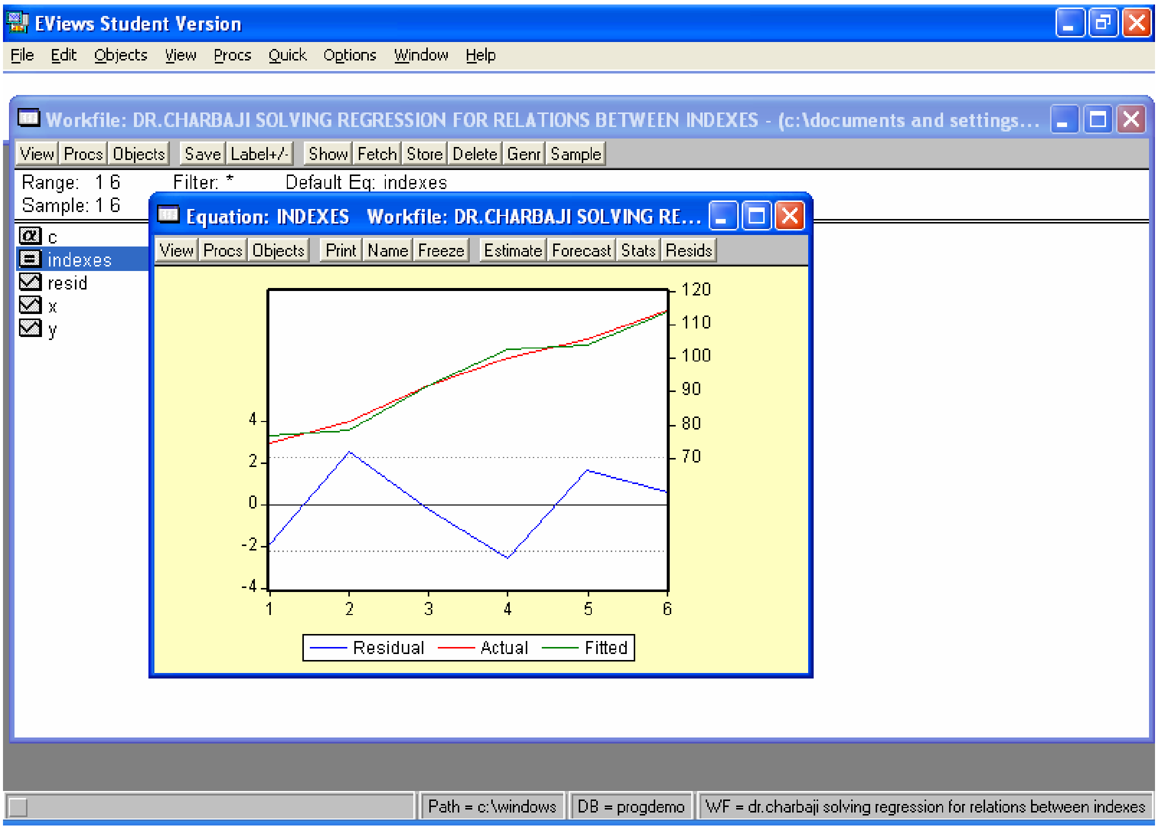
View Representation \ on the Equation (Output) Window



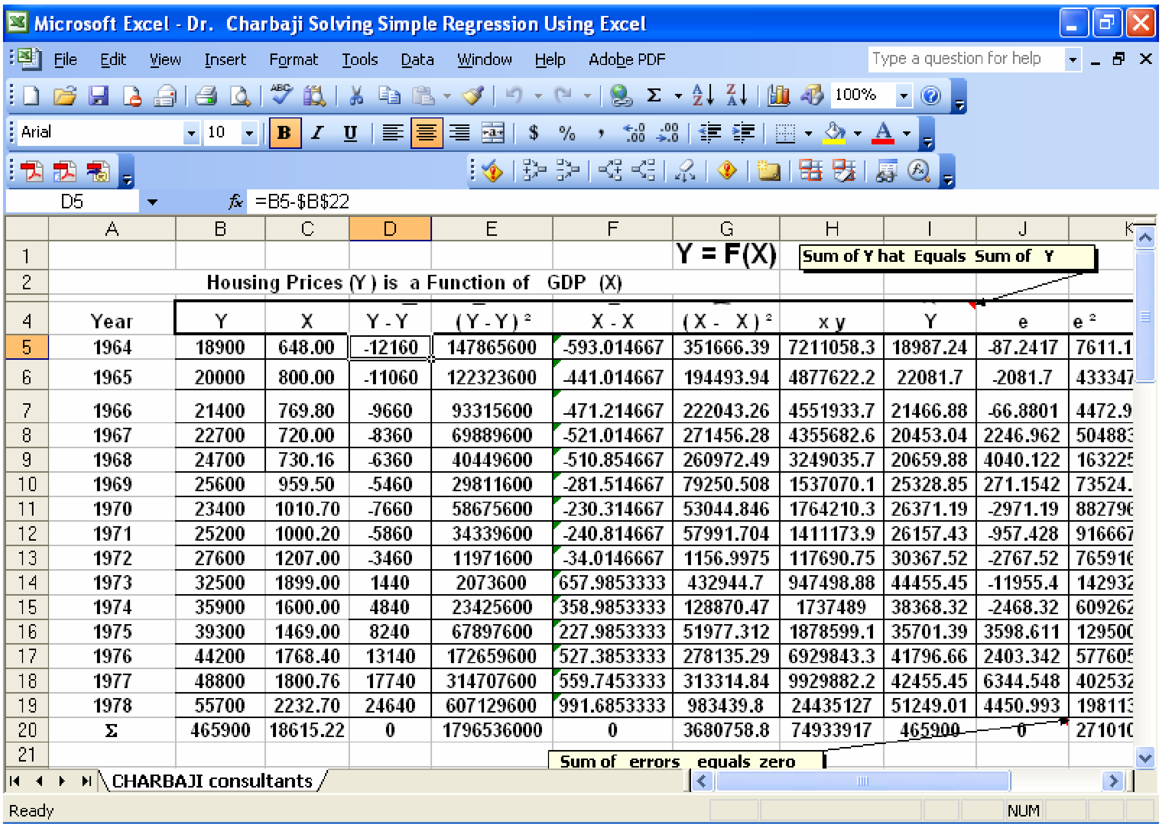
Residual Table \ on the Equation (Output) Window، Fitted،View Actual



Residual Graph \ on the Equation (Output) Window ، Fitted،View Actual



مثال عن الأنحدار الخطي البسيط باتخدام ال Excel واستخدام ال SPSS



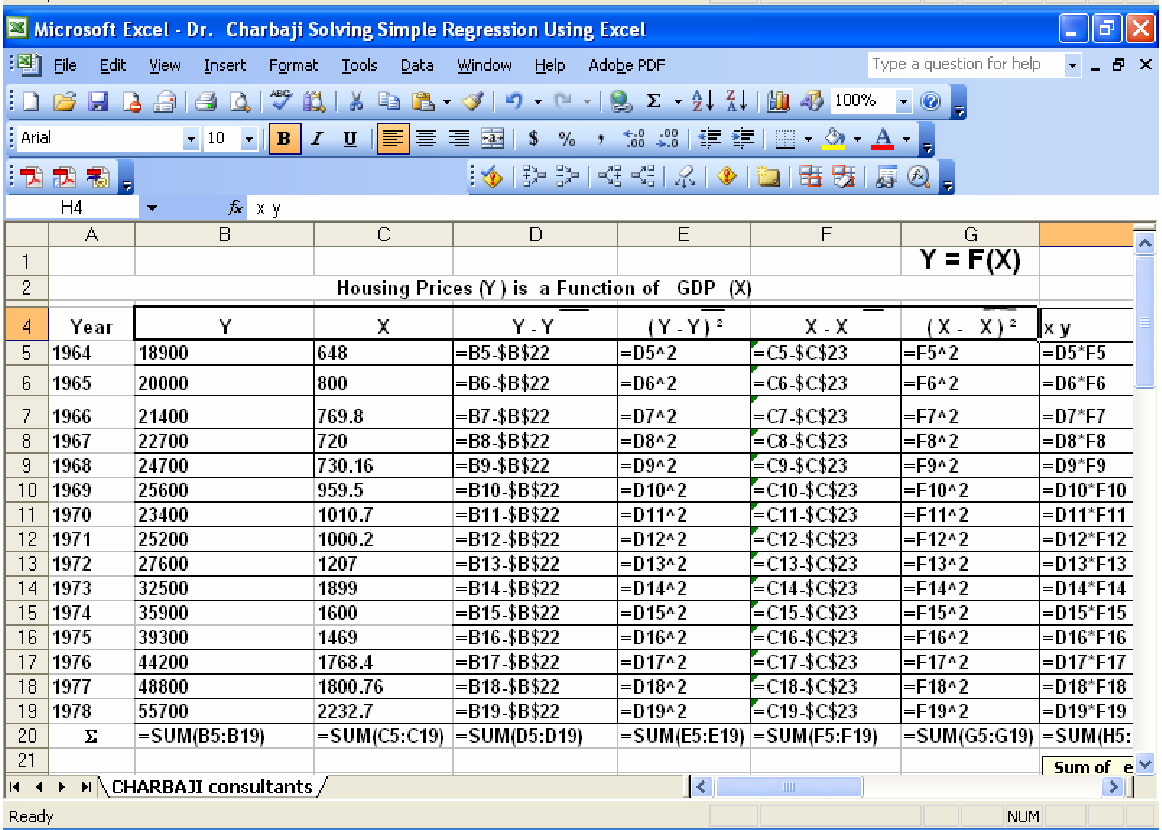
Microsoft Excel - Dr. Charbaji Solving Simple Regression Using Excel

D5 =B5-$B$22

	A	B	C	D	E	F	G	H	I	J	K
1							Y = F(X)	Sum of Y hat Equals Sum of Y			
2		Housing Prices (Y) is a Function of GDP (X)									
4	Year	Y	X	Y - Y	(Y - Y) ²	X - X	(X - X) ²	x y	Y	e	e ²
5	1964	18900	648.00	-12160	147865600	-593.014667	351666.39	7211058.3	18987.24	-87.2417	7611.1
6	1965	20000	800.00	-11060	122323600	-441.014667	194493.94	4877622.2	22081.7	-2081.7	433347
7	1966	21400	769.80	-9660	93315600	-471.214667	222043.26	4551933.7	21466.88	-66.8801	4472.9
8	1967	22700	720.00	-8360	69889600	-521.014667	271456.28	4355682.6	20453.04	2246.962	504883
9	1968	24700	730.16	-6360	40449600	-510.854667	260972.49	3249035.7	20659.88	4040.122	163225
10	1969	25600	959.50	-5460	29811600	-281.514667	79250.508	1537070.1	25328.85	271.1542	73524.
11	1970	23400	1010.70	-7660	58675600	-230.314667	53044.846	1764210.3	26371.19	-2971.19	882796
12	1971	25200	1000.20	-5860	34339600	-240.814667	57991.704	1411173.9	26157.43	-957.428	916667
13	1972	27600	1207.00	-3460	11971600	-34.0146667	1156.9975	117690.75	30367.52	-2767.52	765916
14	1973	32500	1899.00	1440	2073600	657.9853333	432944.7	947498.88	44455.45	-11955.4	142932
15	1974	35900	1600.00	4840	23425600	358.9853333	128870.47	1737489	38368.32	-2468.32	609262
16	1975	39300	1469.00	8240	67897600	227.9853333	51977.312	1878599.1	35701.39	3598.611	129500
17	1976	44200	1768.40	13140	172659600	527.3853333	278135.29	6929843.3	41796.66	2403.342	577605
18	1977	48800	1800.76	17740	314707600	559.7453333	313314.84	9929882.2	42455.45	6344.548	402532
19	1978	55700	2232.70	24640	607129600	991.6853333	983439.8	24435127	51249.01	4450.993	198113
20	Σ	465900	18615.22	0	1796536000	0	3680758.8	74933917	465900	0	271010
21						Sum of errors equals zero					

CHARBAJI consultants

Ready NUM



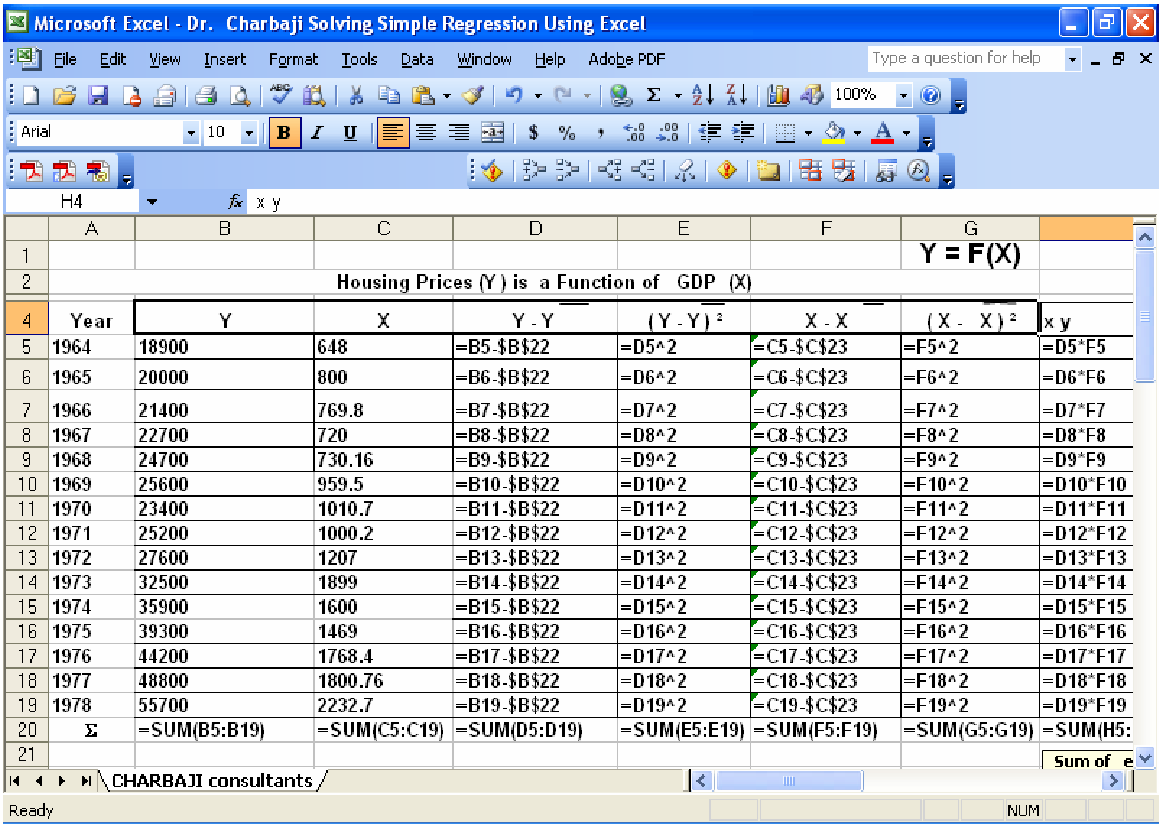
Microsoft Excel - Dr. Charbaji Solving Simple Regression Using Excel

H4 x y

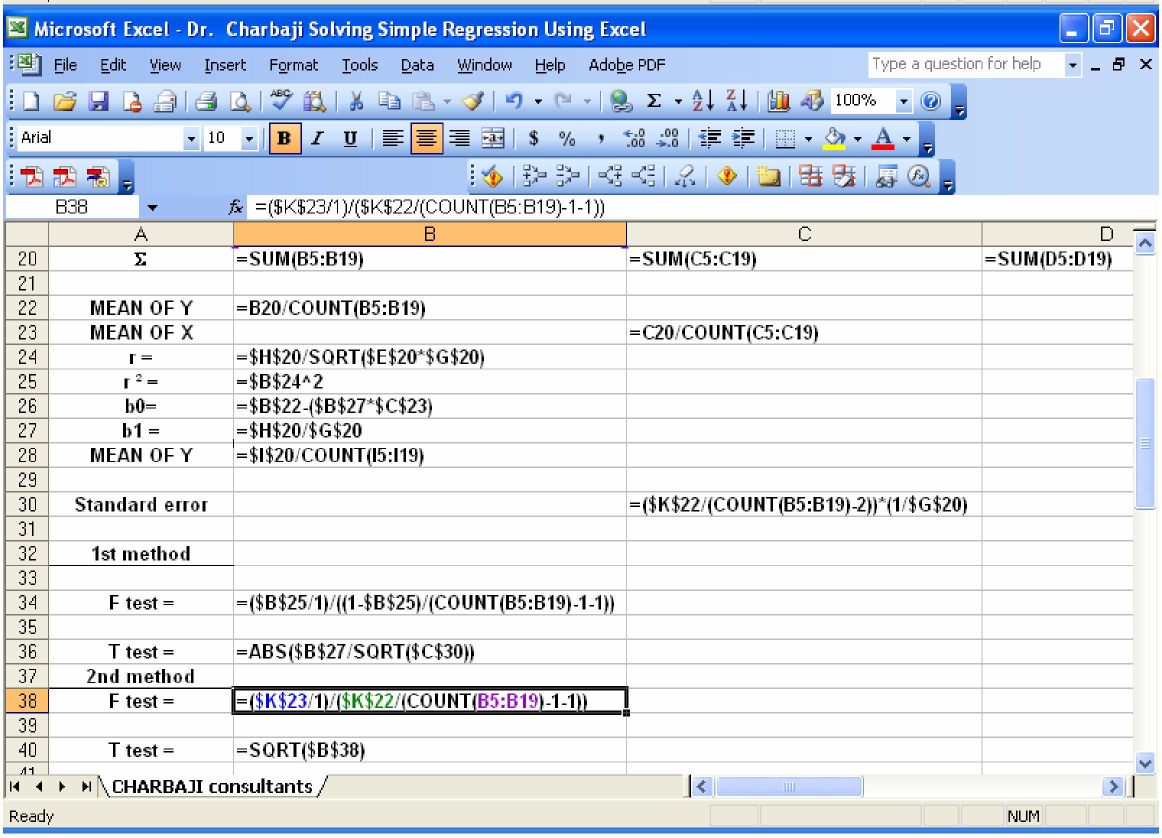
	A	B	C	D	E	F	G	H
1							Y = F(X)	
2			Housing Prices (Y) is a Function of GDP (X)					
4	Year	Y	X	Y - Y	(Y - Y) ²	X - X	(X - X) ²	x y
5	1964	18900	648	=B5-$B$22	=D5^2	=C5-$C$23	=F5^2	=D5*F5
6	1965	20000	800	=B6-$B$22	=D6^2	=C6-$C$23	=F6^2	=D6*F6
7	1966	21400	769.8	=B7-$B$22	=D7^2	=C7-$C$23	=F7^2	=D7*F7
8	1967	22700	720	=B8-$B$22	=D8^2	=C8-$C$23	=F8^2	=D8*F8
9	1968	24700	730.16	=B9-$B$22	=D9^2	=C9-$C$23	=F9^2	=D9*F9
10	1969	25600	959.5	=B10-$B$22	=D10^2	=C10-$C$23	=F10^2	=D10*F10
11	1970	23400	1010.7	=B11-$B$22	=D11^2	=C11-$C$23	=F11^2	=D11*F11
12	1971	25200	1000.2	=B12-$B$22	=D12^2	=C12-$C$23	=F12^2	=D12*F12
13	1972	27600	1207	=B13-$B$22	=D13^2	=C13-$C$23	=F13^2	=D13*F13
14	1973	32500	1899	=B14-$B$22	=D14^2	=C14-$C$23	=F14^2	=D14*F14
15	1974	35900	1600	=B15-$B$22	=D15^2	=C15-$C$23	=F15^2	=D15*F15
16	1975	39300	1469	=B16-$B$22	=D16^2	=C16-$C$23	=F16^2	=D16*F16
17	1976	44200	1768.4	=B17-$B$22	=D17^2	=C17-$C$23	=F17^2	=D17*F17
18	1977	48800	1800.76	=B18-$B$22	=D18^2	=C18-$C$23	=F18^2	=D18*F18
19	1978	55700	2232.7	=B19-$B$22	=D19^2	=C19-$C$23	=F19^2	=D19*F19
20	Σ	=SUM(B5:B19)	=SUM(C5:C19)	=SUM(D5:D19)	=SUM(E5:E19)	=SUM(F5:F19)	=SUM(G5:G19)	=SUM(H5:
21								Sum of e

CHARBAJI consultants

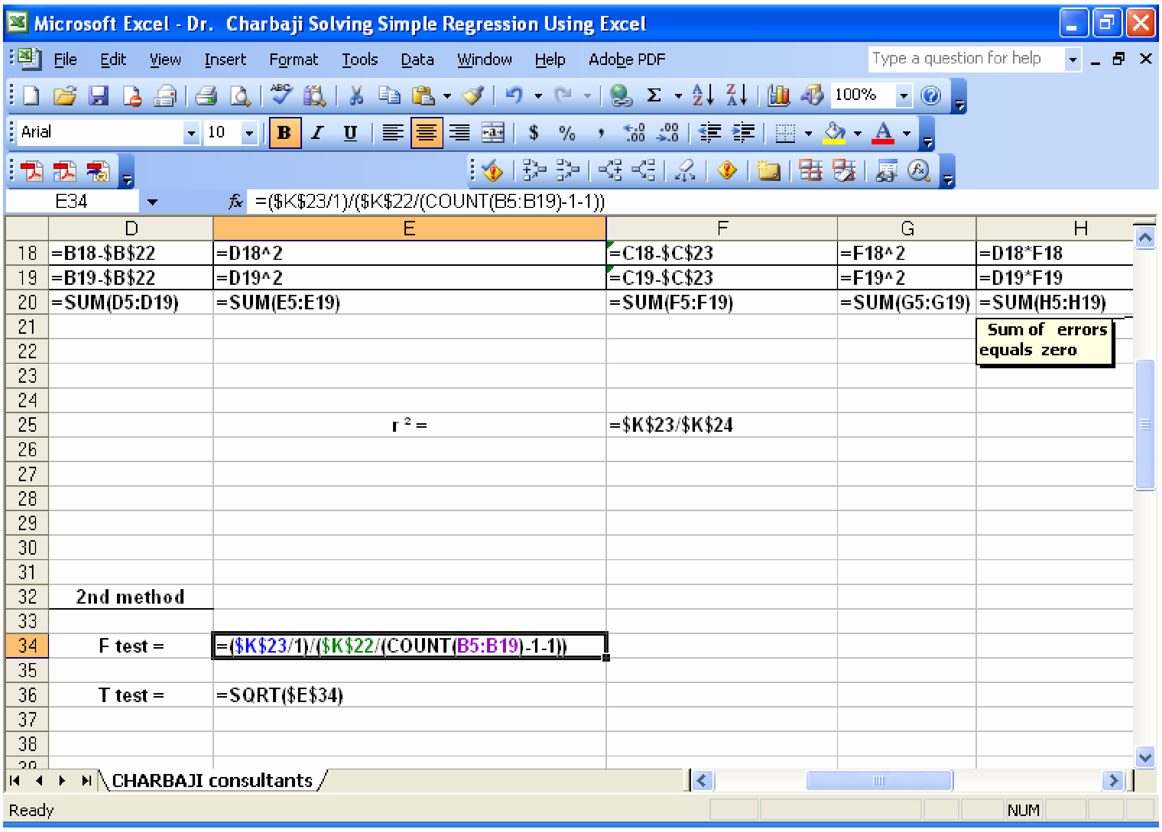
Ready NUM

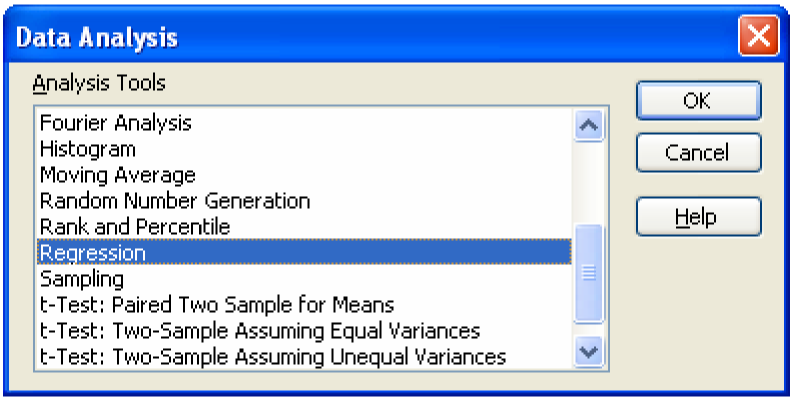


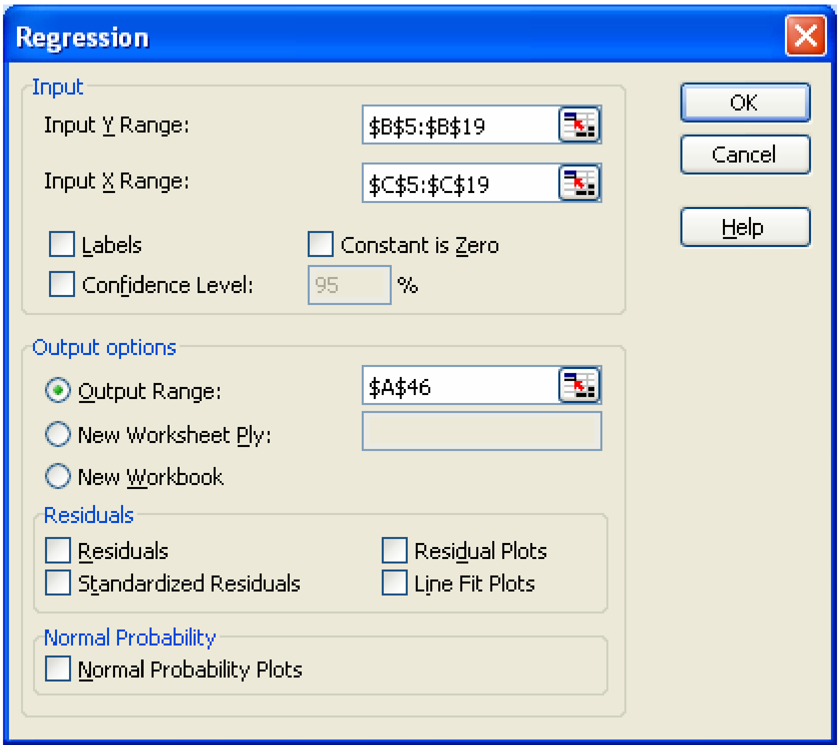
	A	B	C	D	E	F	G	H
1							Y = F(X)	
2		Housing Prices (Y) is a Function of GDP (X)						
4	Year	Y	X	Y - Ȳ	(Y - Ȳ)²	X - X̄	(X - X̄)²	x y
5	1964	18900	648	=B5-$B$22	=D5^2	=C5-$C$23	=F5^2	=D5*F5
6	1965	20000	800	=B6-$B$22	=D6^2	=C6-$C$23	=F6^2	=D6*F6
7	1966	21400	769.8	=B7-$B$22	=D7^2	=C7-$C$23	=F7^2	=D7*F7
8	1967	22700	720	=B8-$B$22	=D8^2	=C8-$C$23	=F8^2	=D8*F8
9	1968	24700	730.16	=B9-$B$22	=D9^2	=C9-$C$23	=F9^2	=D9*F9
10	1969	25600	959.5	=B10-$B$22	=D10^2	=C10-$C$23	=F10^2	=D10*F10
11	1970	23400	1010.7	=B11-$B$22	=D11^2	=C11-$C$23	=F11^2	=D11*F11
12	1971	25200	1000.2	=B12-$B$22	=D12^2	=C12-$C$23	=F12^2	=D12*F12
13	1972	27600	1207	=B13-$B$22	=D13^2	=C13-$C$23	=F13^2	=D13*F13
14	1973	32500	1899	=B14-$B$22	=D14^2	=C14-$C$23	=F14^2	=D14*F14
15	1974	35900	1600	=B15-$B$22	=D15^2	=C15-$C$23	=F15^2	=D15*F15
16	1975	39300	1469	=B16-$B$22	=D16^2	=C16-$C$23	=F16^2	=D16*F16
17	1976	44200	1768.4	=B17-$B$22	=D17^2	=C17-$C$23	=F17^2	=D17*F17
18	1977	48800	1800.76	=B18-$B$22	=D18^2	=C18-$C$23	=F18^2	=D18*F18
19	1978	55700	2232.7	=B19-$B$22	=D19^2	=C19-$C$23	=F19^2	=D19*F19
20	Σ	=SUM(B5:B19)	=SUM(C5:C19)	=SUM(D5:D19)	=SUM(E5:E19)	=SUM(F5:F19)	=SUM(G5:G19)	=SUM(H5:
21								Sum of e

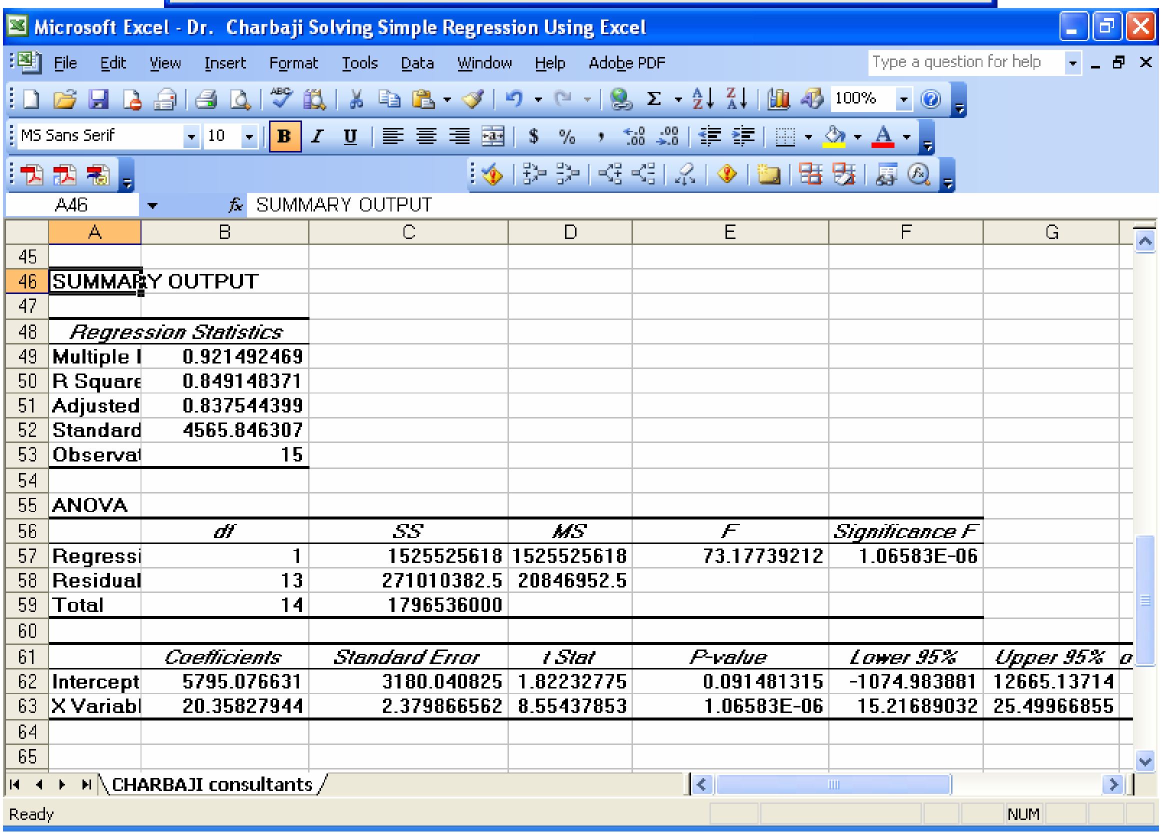


	A	B	C	D
20	Σ	=SUM(B5:B19)	=SUM(C5:C19)	=SUM(D5:D19)
21				
22	MEAN OF Y	=B20/COUNT(B5:B19)		
23	MEAN OF X		=C20/COUNT(C5:C19)	
24	r =	=$H$20/SQRT($E$20*$G$20)		
25	r ² =	=$B$24^2		
26	b0=	=$B$22-($B$27*$C$23)		
27	b1 =	=$H$20/$G$20		
28	MEAN OF Y	=$I$20/COUNT(I5:I19)		
29				
30	Standard error		=($K$22/(COUNT(B5:B19)-2))*(1/$G$20)	
31				
32	1st method			
33				
34	F test =	=($B$25/1)/((1-$B$25)/(COUNT(B5:B19)-1-1))		
35				
36	T test =	=ABS($B$27/SQRT($C$30))		
37	2nd method			
38	F test =	=($K$23/1)/($K$22/(COUNT(B5:B19)-1-1))		
39				
40	T test =	=SQRT($B$38)		


Microsoft Excel - Dr. Charbaji Solving Simple Regression Using Excel
E34
=($K$23/1)/($K$22/(COUNT(B5:B19)-1-1))
=B18-$B$22
=D18^2
=C18-$C$23
=F18^2
=D18*F18
=B19-$B$22
=D19^2
=C19-$C$23
=F19^2
=D19*F19
=SUM(D5:D19)
=SUM(E5:E19)
=SUM(F5:F19)
=SUM(G5:G19)
=SUM(H5:H19)
Sum of errors equals zero
r ² =
=$K$23/$K$24
2nd method
F test =
=($K$23/1)/($K$22/(COUNT(B5:B19)-1-1))
T test =
=SQRT($E$34)
CHARBAJI consultants
Ready
NUM


Data Analysis
Analysis Tools
Fourier Analysis
Histogram
Moving Average
Random Number Generation
Rank and Percentile
Regression
Sampling
t-Test: Paired Two Sample for Means
t-Test: Two-Sample Assuming Equal Variances
t-Test: Two-Sample Assuming Unequal Variances
OK
Cancel
Help





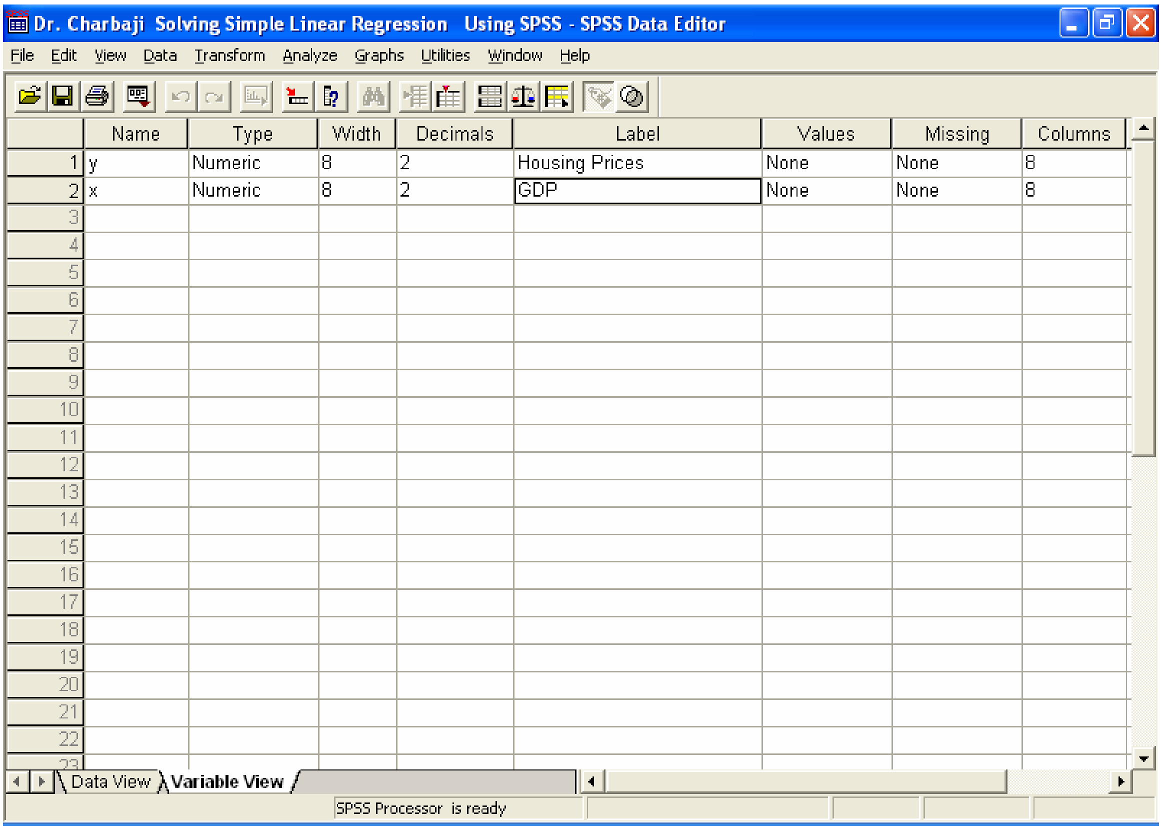
SUMMARY OUTPUT

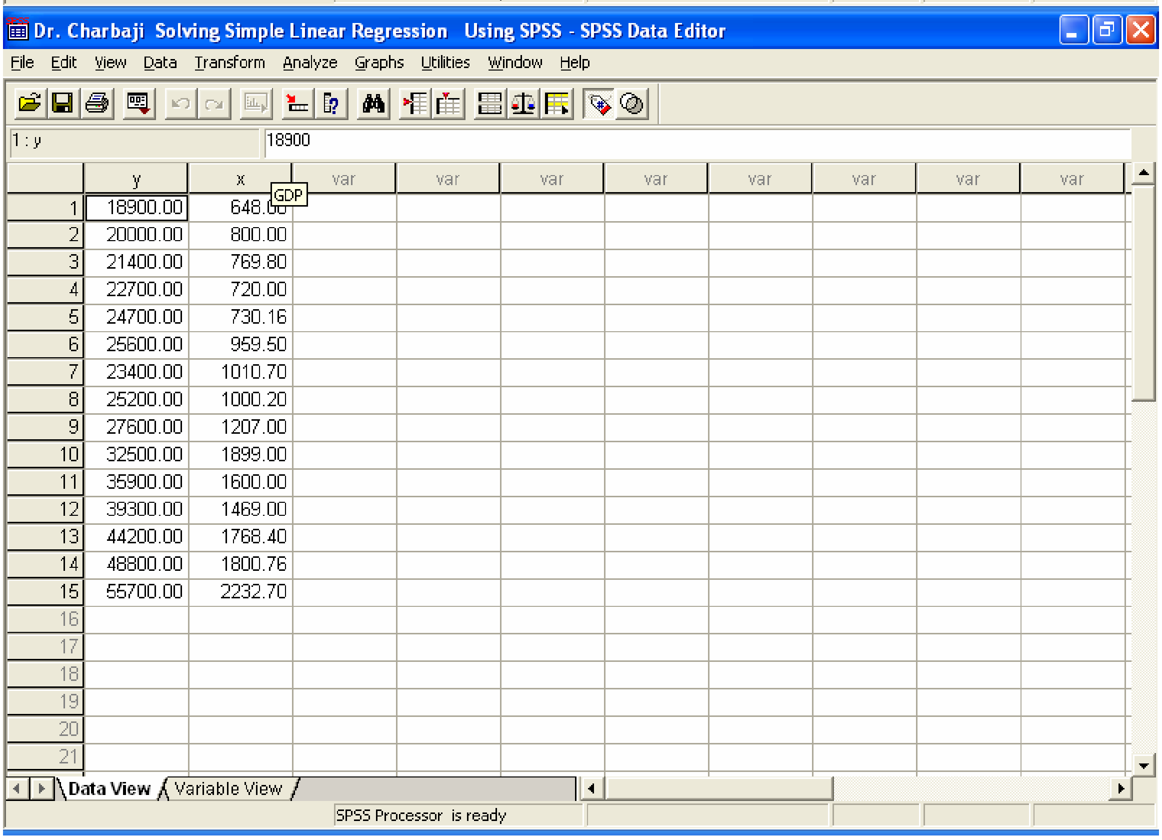
Regression Statistics	
Multiple R	0.921492469
R Square	0.849148371
Adjusted	0.837544399
Standard	4565.846307
Observat	15

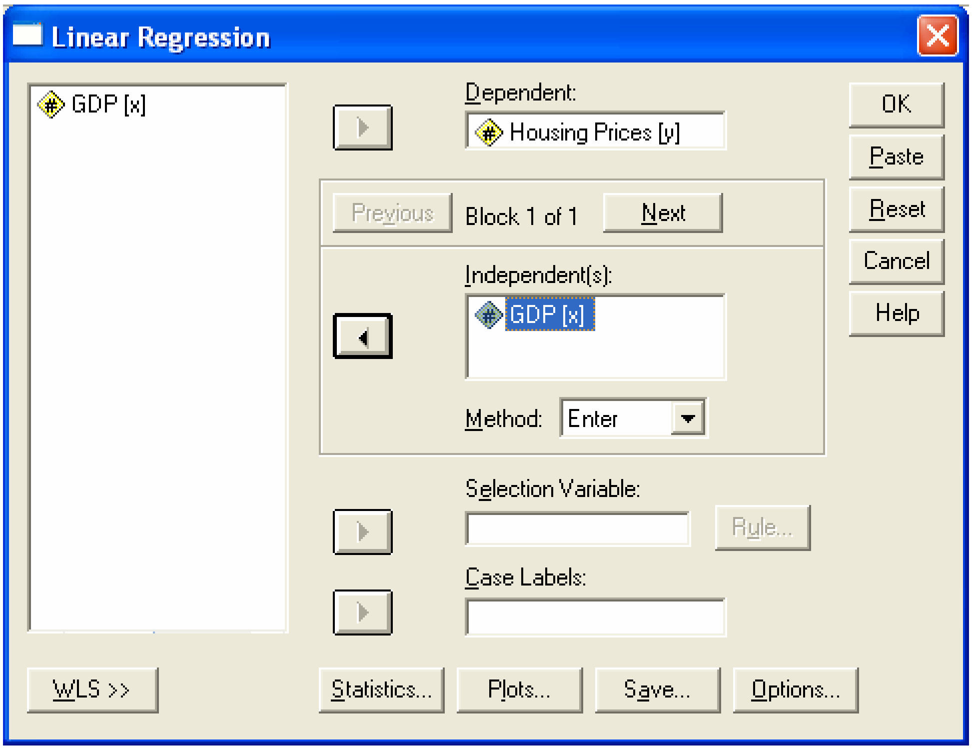
ANOVA

	df	SS	MS	F	Significance F
Regressi	1	1525525618	1525525618	73.17739212	1.06583E-06
Residual	13	271010382.5	20846952.5		
Total	14	1796536000			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	5795.076631	3180.040825	1.82232775	0.091481315	-1074.983881	12665.13714
X Variabl	20.35827944	2.379866562	8.55437853	1.06583E-06	15.21689032	25.49966855


Dr. Charbaji Solving Simple Linear Regression Using SPSS - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Name Type Width Decimals Label Values Missing Columns
1 y Numeric 8 2 Housing Prices None None 8
2 x Numeric 8 2 GDP None None 8
Data View Variable View
SPSS Processor is ready


Dr. Charbaji Solving Simple Linear Regression Using SPSS - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
1 : y 18900
y x var var var var var var var var
GDP
1 18900.00 648.00
2 20000.00 800.00
3 21400.00 769.80
4 22700.00 720.00
5 24700.00 730.16
6 25600.00 959.50
7 23400.00 1010.70
8 25200.00 1000.20
9 27600.00 1207.00
10 32500.00 1899.00
11 35900.00 1600.00
12 39300.00 1469.00
13 44200.00 1768.40
14 48800.00 1800.76
15 55700.00 2232.70
Data View Variable View
SPSS Processor is ready



Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.921[a]	.849	.838	4565.84631

a. Predictors: (Constant), GDP

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.53E+09	1	1525525618	73.177	.000[a]
	Residual	2.71E+08	13	20846952.50		
	Total	1.80E+09	14			

a. Predictors: (Constant), GDP

b. Dependent Variable: Housing Prices

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5795.077	3180.041		1.822	.091
	GDP	20.358	2.380	.921	8.554	.000

a. Dependent Variable: Housing Prices

Exercises

1. The following table gives data on weekly consumption expenditure (Y) (in dollars) and weekly family income (X) (in dollars). (Le tableau suivant présente des données sur la consommation hebdomadaire (Y) (en dollars) et le revenu familial hebdomadaire (X) (en dollars)).

Table 3: Income & consumption

X ($)	Y ($)
80	70
100	65
120	90
140	95
160	110
180	115
200	120
220	140
240	155
260	150

   a) Is this a time series regression or a cross-sectional regression? (est-ce une série chronologique ou une série en coupe instantanée ?)
   b) Use the least squares method pour regress Y on X. (Utiliser la méthode des moindres carrés pour régresser Y sur X)
   c) How do you interpret the parameters of this regression? (Comment interpréter les paramètres de cette régression)
   d) Test the significance of these parameters, α = 0.05. (tester la signification de ces paramètres, α = 0.05)
   e) Find the coefficient of determination; interpret. (Trouver le coefficient de détermination ; interpréter)

2. A store manager selling TV sets observes the following sales on 10ddifferent days. (Un directeur de magasin qui vend des télévisions a observé les ventes sur 10 jours différents)
   Y= Number of TV sets = Nombre de télévisions
   X=Number of sales representatives = Nombre de vendeurs

Table 2: Number of sets and sales representatives

Y	**3**	**6**	**10**	**5**	**10**	**12**	**5**	**10**	**10**	**8**
X	**1**	**1**	**1**	**2**	**2**	**2**	**3**	**3**	**3**	**2**

   a) Plot the data with Y on the vertical axis and X on the horizontal axis. (Représenter graphiquement les données avec Y en ordonnées et X en abscisses)
   b) What can you say about the relationship between these two variables? (Comment peut-on dire de la relation entre ces deux variables?)
   c) Calculate the regression of Y on X. (Calculer la régression de Y sur X)
   d) Test the significance of the parameters, α = 0.05. Interpret the results. (tester la signification des paramètres, α = 0.05. Interpréter les résultats)
   e) Establish a 95% confidence interval for the true parameters. (Etablir l'intervalle de confiance à 95% pour les paramètres de la régression)

3. In the model $Y_t = \alpha + \beta X_t + u_t$, i =1,.., N, the following sample moments have been calculated from 10 observations: (Dans le modèle $Y_t = \alpha + \beta X_t + u_t$, i =1, ... , N, les moments suivant d'un échantillon de 10 observations ont été calculés)

   $\sum Y_t = 8$ $\quad \sum X_t = 40$ $\quad \sum X_t Y_t = 20$ $\quad \sum X_t^2 = 200$ $\quad \sum Y_t^2 = 26$

   a) Calculate the predictor of Y for X = 10 and obtain a 95% confidence interval for it. (Prédire la valeur de Y sachant que X = 10 et obtenir son intervalle de confiance à 95%)
   b) Calculate the value of X that could have given rise to a value of Y = 1 and explain how you would find a 95% confidence interval for it. (Calculer la valeur de X qu'aurait pu engendrer une valeur d'Y = 1 et expliquer comment vous trouveriez son intervalle de confiance à 95%)

4. In the following table we have data on CPI and S&P 500 Index for the years 1990 to 2002. (Le tableau suivant fournit des données annuelles sur l'indice des prix à consommation (CPI) et L'indice S&P 500 sur la période 1990 – 2002)

Table 4: S&P and CPI (1990-2002)

Year (Année)	S&P	CPI
1990	334.59	130.7
1991	376.18	136.2
1992	415.74	140.3
1993	451.41	144.5
1994	460.42	148.2
1995	541.72	152.4
1996	670.50	156.9
1997	873.43	160.5
1998	1085.50	163.0
1999	1327.33	166.6
2000	1427.22	172.2
2001	1194.18	177.1

a) Plot the data on a scattergram with the S&P 500 on the vertical axis and IPC on the horizontal axis. (Représenter graphiquement l'S&P 500 sur l'axe des ordonnées et l'IPC sur celui des abscisses)

b) What can you say about the relationship between the two indexes? What does economic theory have to say about the relationship? (Que peut-on dire de la relation entre ces deux indices? Que dit la théorie économique sur cette relation?)

Consider the following regression model. (Considérer le modèle de régression suivant:)

$$(S\&P)_t = \beta_1 + \beta_2\, CPI_t + u_t$$

Use the least squares to estimate this equation and interpret the results. (Utiliser les moindres carrés pour estimer cette équation et interpréter les résultats)

5. The following table Gives data on the gold price, CPI, and New York Stock Exchange (NYSE) over the period 1977-1991. (Le tableau suivant présente des données sur le prix de l'or, l'indice des prix à la consommation et l'indice du New York Stock Exchange (NYSE) pour la période 1977-1991)

Table 5: NYSE, CPI and Gold price (1977-1991)

NY stock Exchange (NYSE) 31 Dec.1965 = 100	CPI 1982 – 1984 = 100	Gold price in \$ per once (31.1g)	Year (Année)
53.69	60.6	147.98	1977
53.70	65.2	193.44	1978
58.32	72.6	307.62	1979
68.10	82.4	612.51	1980
74.02	90.9	459.61	1981
68.93	96.5	376.01	1982
92.63	99.6	423.83	1983
92.46	103.9	360.29	1984
108.90	107.6	317.3	1985
136.00	109.6	367.87	1986
161.70	113.6	446.5	1987
149.91	118.3	436.93	1988
180.02	124.0	381.28	1989
183.46	130.7	384.04	1990
206.33	136.2	362.04	1991

a) Plot on the same scattergram the Gold prices, CPI, and NYSE index. (Représenter sur le même graphique les prix de l'or, l'indice des prix et l'indice du NYSE)

b) An investment is supposed to be a protection against inflation. To test this hypothesis, suppose that you decide to adjust the following model, given that the scattergram in (a) is well fitting the data (Un investissement est supposé être une protection contre l'inflation. Pour tester cette hypothèse, supposez que vous décidez d'ajuster le modèle suivant, en supposant que le graphique de dispersion du (a) est approprié)

$$Gold\ price_t = \beta_1 + \beta_2 CPI_t + u_t$$

$$NYSE\ index_t = \beta_1 + \beta_2 CPI_t + u_t$$

الفصل الثالث
الانحدار الخطي المتعدّد

تمهيد:

لقد تم في الفصل الثاني مناقشة الانحدار الخطي البسيط لمتغير تابع على متغير مستقل واحد. وبما أن واقع الحياة الاقتصادية ، يقوم بشكل عام على تأثر أية ظاهرة، بأكثر من متغير مستقل ، لذلك لا بد من توسيع الأسلوب الذي مرّ معنا في الفصل السابق ليشتمل على انحدار المتغير التابع على العديد من المتغيرات المستقلة. علماً أن توافر الآلات الحاسبة الالكترونية، هو الذي مكن الباحثين من اجتياز صعوبة العمليات الحسابية في تحليل الانحدار المتعدد ومعالجة النماذج الاقتصادية التي تتضمن العديد من المعادلات والعديد من المتغيرات في آن واحد (Simultaneous equation models – Modèles à équations simultanées). سنكتفي في هذا الفصل بمناقشة الانحدار الخطي المتعدد (Multiple linear regression - Régression linéaire multiple) في حين نناقش موضوع تحليل المتغيرات المتعددة (Multivariate analysis – Analyse multivariée) للنماذج الآنية (Simultaneous models – Modèles simultanés) في فصول قادمة. علماً أن تحليل الانحدار المتعدد هو التحليل الذي يُستخدم في اختيار الفروض حول العلاقة بين متغير تابع وبين متغير أو أكثر من المتغيرات المستقلة.

معاملات الانحدار الجزئية:

تعرف معاملات الانحدار، في تحليل الانحدار المتعدد، باسم معاملات الانحدار الجزئية (Partial regression coefficient – Coefficient de regression partielle)، وذلك تمييزاً عن معامل الانحدار البسيط الذي مرّ معنا في الفصل السابق. لقد سبق وذكرنا أن معامل الانحدار البسيط b_1 في تحليل الانحدار الخطي البسيط، يقيس معدل التغير المتوقع في المتغير التابع Y نتيجة تغير المتغير المستقل X_1 بوحدة قياس واحدة. أما معامل الانحدار الجزئي b_i في تحليل الانحدار المتعدد، فيقيس معدل التغير المتوقع في المتغير التابع Y، نتيجة تغير المتغير المستقل X_i بوحدة قياس واحدة، مع بقاء أثر بقية المتغيرات المستقلة الأخرى ثابتاً (Holding the effect of other independent variables constant – En gardant les autres variables indépendantes constantes). وتكتب المعادلة الحقيقية لانحدار المتغير Y على المتغيرين X_1 و X_2 كالآتي:

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

بينما تكتب المعادلة الحقيقية لانحدار Y على K من المتغيرات المستقلة كالآتي:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_K X_K + U$$

وسيتم التركيز في هذا الفصل على انحدار Y على متغيرين مستقلين X_1 و X_2. علماً أنه يمكننا استعمال ذات الأسلوب المستخدم في تحليل انحدار Y على متغيرين، في تحليل انحدار Y على K من المتغيرات المستقلة. وتجدر الإشارة هنا إلى أن الفروض الخاصة بالخطأ العشوائي

أولاً: الحصول على معاملات الانحدار الجزئيةٍ باستخدام المعادلات الطبيعية لمستوى انحدار المربعات الصغرى.

ثانياً: الحصول على معاملات الانحدار الجزئية باستخدام المصفوفات للوحدات الخام.

ثالثاً: الحصول على معاملات الانحدار الجزئية باستخدام المصفوفة معاملات الارتباط بين المتغيرات .

ثالثاً : الحصول على معاملات الانحدار الجزئية باستخدام البواقي .

إذن: سنتناول كلٍ من هذه الطرق بالتفصيل:

أولاً: الحصول على معاملات الانحدار الجزئية باستخدام المعادلات الطبيعية:

يمكن صياغة المعادلة التقديرية لانحدار المتغير التابع Y، على المتغيرين المستقلين X_1 و X_2، كالآتي:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + e$$

وتتلخص المشكلة، في الحصول على معاملات الانحدار الجزئية b_0, b_1, b_2 والتي تجعل مجموع مربع البواقي $\sum e^2 = \sum(Y - \acute{Y})^2$ عند نهايتها الصغرى. علماً أنه لتبسيط العمليات الحسابية، فباستطاعتنا استخدام المتغيرات في شكل انحرافات عن الأوساط الحسابية فتصبح المعادلة التقديرية للانحدار كالآتي:

$$\acute{y} = b_1 x_1 + b_2 x_2 \qquad (1)$$

وللحصول على المعادلات الطبيعية (Normal equations – Les équations normales)، بطريقة سهلة، ودون استعمال التفاضل الجزئي(Partial derivatives)، فإننا نضرب (Multiply - Multiplier) طرفي المعادلة (1)، بالمتغير x_1، ونجمع طرفي المعادلة فنحصل على:

$$\sum x_1 y = b_1 \sum x_1^2 + b_2 \sum x_1 x_2 \qquad (2)$$

ثم نضرب طرفي المعادلة (1)، بالمتغير x_2 ، ونجمع طرفي المعادلة فنحصل على:

$$\sum x_2 y = b_1 \sum x_1 x_2 + b_2 \sum x_2^2 \qquad (3)$$

وباستخدام طريقة كرايمر يتم الحصول على b_1 و b_2 ، من المعادلتين (2) و (3) كالآتي:*

$$b_1 = \frac{\begin{vmatrix} \sum x_1 y & \sum x_1 x_2 \\ \sum x_2 y & \sum x_2^2 \end{vmatrix}}{\begin{vmatrix} \sum x_1^2 & \sum x_1 x_2 \\ \sum x_1 x_2 & \sum x_2^2 \end{vmatrix}} = \frac{(\sum x_1 y)(\sum x_2^2) - (\sum x_2 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

$$b_2 = \frac{\begin{vmatrix} \sum x_1^2 & \sum x_1 y \\ \sum x_1 x_2 & \sum x_2 y \end{vmatrix}}{\begin{vmatrix} \sum x_1^2 & \sum x_1 x_2 \\ \sum x_1 x_2 & \sum x_2^2 \end{vmatrix}} = \frac{(\sum x_2 y)(\sum x_1^2) - (\sum x_1 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

* لاحظ أنه يتم الحصول على المعادلات الطبيعية عادة باستخدام التفاضل الجزئي (partial derivatives – dérivés partielles) بالنسبة للمعامل b_1 و b_2 , وجعل المفاضلة الجزئية مساوية للصفر كالآتي:

$$\Sigma e^2 = \Sigma(y - b_1 x_1 - b_2 x_2)^2$$

$$\frac{\partial \Sigma e^2}{\partial b_1} = -2\Sigma(y - b_1 x_1 - b_2 x_2)(x_1) = 0$$

$$\frac{\partial e^2}{\partial b_2} = -2\Sigma(y - b_1 x_1 - b_2 x_2)(x_2) = 0$$

أما الثابت b_0 فيتم الحصول عليه بجمع طرفي معادلة الانحدار

$Y = b_0 + b_1 X_1 + b_2 X_2$ وقسمة الناتج على n كالآتي :

$$\frac{\sum Y}{n} = \frac{b_0 n}{n} + \frac{b_1 \sum X_1}{n} + \frac{b_2 \sum X_2}{n}$$

$$\bar{Y} = b_0 + b_1 \bar{X}_1 + b_2 \bar{X}_2$$

إذن:

$$b_0 = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2$$

أما تباين توزيع المعانية لمعامل الانحدار الجزئي، والذي يستخدم في اختبار الفروض وتكوين فترات الثقة فهو:

$$S^2_{b_1} = \frac{SS_{\mathrm{Resd}}}{n-k-1} \cdot \frac{\sum x_2^2}{\left(\sum x_1^2\right)\left(\sum x_2^2\right) - \left(\sum x_1 x_2\right)^2}$$

$$S^2_{b_2} = \frac{SS_{\mathrm{Resd}}}{n-k-1} \cdot \frac{\sum x_1^2}{\left(\sum x_1^2\right)\left(\sum x_2^2\right) - \left(\sum x_1 x_2\right)^2}$$

ويصاغ فرض العدم (H_0) والفرض البديل (H_1) في حالة الانحدار المتعدد كالآتي :

فرض العدم : كل معاملات الانحدار الجزئية في المجتمع الإحصائي تساوي صفراً .

$$H_0 = \beta_1 = \beta_2 = \beta_3 \ldots\ldots = \beta = 0$$

الفرض البديل: على الأقل فإنّ إحدى معاملات الانحدار الجزئية في المجتمع الإحصائي لا تساوي صفراً. (بمعنى أنه على الأقل يوجد متغيّر مستقل-من بين المتغيّرات المستقلة – جوهري من الناحية الإحصائية).

ويمكن اختبار جوهرية معامل الانحدار الجزئي باستخدام اختبار t الإحصائي، كالآتي:

$$t = \frac{b_i - \beta}{S_{b_i}}$$

كذلك يمكننا صياغة فترة الثقة لمعامل الانحدار الجزئي:

$$\beta = b_i \; C \, t_{(0.05,\, n-k-1)} S_{b_i}$$

وأخيراً تجدر الإشارة إلى ضرورة الانتباه في تفسير معاملات الانحدار الجزئية، في حالة استخدام بيانات لسلاسل زمنية (Time series data – Données chronologiques) وفي حالة استخدام بيانات عن متغيرات مأخوذة في نقطة زمنية محددة (Cross-section data – Données en coupes instantanées).

دعنا نفترض أنّ أحد الباحثين، يرغب في تحديد الاختلافات في الإنفاق السنوي الاستهلاكي، على السلعة Z (مُقاساً بالمائة ليرة)، لعينة من العائلات، من خلال معرفته بالاختلافات على متغيرين مستقلين: X_1 (الدخل السنوي للعائلات مقاساً بالألف ليرة) و X_2 (حجم العائلة مقاساً بعدد أفراد الأسرة)، وأنّه حصل على معادلة الانحدار الآتية:

$$\hat{Y} = 4.20 + 3.49\,X_1 - 0.57\,X_2$$

فحينئذٍ باستطاعته تفسير b_1، على أنّه إذا كان لدينا عائلتين بنفس الحجم (عدد أفراد الأسرة)، بحيث يزيد دخل العائلة الأولى، على دخل العائلة الثانية، بمقدار وحدة قياس واحدة (ألف ليرة)، فإنّ الإنفاق الاستهلاكي التقديري للعائلة الأولى على السلعة Z، سوف يزيد عن إنفاق العائلة الثانية على السلعة Z، بمقدار 3.49 من وحدة القياس للمتغيّر التابع (مائة ليرة). أي أنّه إذا كان لدينا عائلتين **بنفس الحجم**، بحيث يزيد دخل إحدى العائلتين على دخل العائلة الأخرى، بمقدار ألف ليرة، فإنّ الإنفاق الاستهلاكي، على السلعة Z، للعائلة الأولى سيزيد على الإنفاق الاستهلاكي على السلعة Z للعائلة الثانية بمقدار 349 ليرة. ويمكننا تفسير b_2 على أنّه إذا كان لدينا عائلتين، **متساويتين في الدخل السنوي**، لكن يزيد عدد أفراد العائلة الثانية على عدد أفراد العائلة الأولى بشخص واحد، فإنّ الإنفاق الاستهلاكي على السلعة Z للعائلة الثانية، سيكون أقل من الإنفاق الاستهلاكي على السلعة Z للعائلة الأولى ، بمقدار 57 ليرة. أي باختصار فإنّ معامل الانحدار الجزئي يقيس أثر التغيّر في X_i على Y، مع بقاء أثر بقية المتغيّرات المستقلة الأخرى ثابتاً (Cetris paribus – Toutes choses égales par ailleurs).

لنفرض الآن أنّ باحثاً آخراً رغب في دراسة أثر أسعار الأسهم العادية مُقاساً بالرقم القياسي للأسهم العادية (Price index of common stocks – Indice des prix des actions ordinaires) بالفترة الحالية، وأثر الأرباح غير الموزعة (Retained earnings – Gains retenus) بالفترة الحالية، مُقاساً بالمليون ليرة، وأثر سعر الفائدة (The interest rate at which funds could be borrowed – Le taux d'intérêt auquel les fonds peuvent être empruntés) على حجم الاستثمار في الفترة اللاحقة (The level of investment – Le niveau d'investissement)، مُقاساً بالمليون ليرة، فجمع بيانات فصلية (Quarterly data – Données trimestrielles) وحصل على معادلة الانحدار التالية:

$$\hat{Y} = 2.3 + 0.091\,X_1 + 1.87\,X_2 + 0.02\,X_3$$

فحينئذٍ يمكنه تفسير معامل الانحدار الجزئي b_1، على أنّه إذا تغيّر الرقم القياسي لأسعار الأسهم العادية بنقطة واحدة (One-point change – Un point de changement)، مع بقاء أثر بقية المتغيّرات الأخرى ثابتاً (Cetris paribus - Toutes choses égales par ailleurs)، فإنّ الاستثمار في الفترة اللاحقة سيزيد بمقدار 0.091 من المليون ليرة. وبنفس المنطق يمكننا تفسير b_2 وb_3 على أنها معدل التغيّر في Y نتيجة تغيّر X_i بوحدة قياس واحدة، مع بقاء أثر بقية المتغيّرات المستقلة الأخرى ثابتاً.

ثانياً: الحصول على معاملات الانحدار الجزئية باستخدام المصفوفات للوحدات الخام:[23]

باستطاعة الباحث بدلاً من استخدام المعادلات الطبيعية في الحصول على معاملات الانحدار الجزئية، أن يلجأ إلى استخدام المصفوفات (Matrices - Matrices) لهذا الغرض، علماً أنّ طريقة المصفوفات، هي الطريقة المستخدمة في برامج الكمبيوتر (Computer software - Logiciels) لتحليل الانحدار. فلو فرضنا على سبيل المثال، أنّنا نرغب في تحليل انحدار المتغيّر التابع Y على المتغيريّن المستقلين X_1 و X_2، فيمكننا وقتئذٍ صياغة معادلة الانحدار بالوحدات الخام كالآتي:

$$Y = b_0 + b_1 X_1 + b_2 X_2 \qquad (1)$$

ويجمع طرفي المعادلة (1) نحصل على:

$$\sum Y = b_0 n + b_1 \sum X_1 + b_2 \sum X_2 \qquad (2)$$

وبضرب طرفي المعادلة (1) بالمتغيّر X_1 وجمع طرفي المعادلة نحصل:

$$\sum X_1 Y = b_0 \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2 \qquad (3)$$

وبضرب طرفي المعادلة (1) بالمتغيّر X_2 وجمع طرفي المعادلة نحصل على:

$$\sum X_2 Y = b_0 \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2 \qquad (4)$$

ويمكن صياغة المعادلات الطبيعية (2) و (3) و (4) في شكل مصفوفات كالآتي:

$$(X'X) \qquad\qquad b \qquad\qquad (X'Y)$$

[23] يقصد بالوحدات الخام (Raw scores) القياسات الأصلية (Original scores) على المتغير والتي لم تخضع بعد إلى أية معالجة إحصائية (Statistical treatment).

$$\begin{bmatrix} n & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1^2 & \sum X_1X_2 \\ \sum X_2 & \sum X_1X_2 & \sum X_2^2 \end{bmatrix} * \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum X_1Y \\ \sum X_2Y \end{bmatrix}$$

علماً أنّ المصفوفة الأولى، إلى أقصى الشمال، هي في الحقيقة عبارة عن حاصل ضرب المصفوفة التي تحتوي على المتغيرات المستقلة X بالمصفوفة المحورة (Transpose – Transposé) X' في حين أنّ الموجه (Vector - Vecteur) إلى أقصى اليمين هو في الحقيقة عبارة عن حاصل ضرب المصفوفة المحورة X' بالموجه Y.

دعنا نفترض للتوضيح أنه لدينا البيانات الفرضية التالية لأربعة مشاهدات (Observations - Observations) على المتغيرات Y و X_1 و X_2:

X_1	X_2	Y
X_{11}	X_{12}	Y_1
X_{21}	X_{22}	Y_2
X_{31}	X_{32}	Y_3
X_{41}	X_{42}	Y_4

حيث ترمز الرموز السفلية (Subscripts - Indices) إلى المشاهدة (Observation - Observation) وإلى المتغيّر على التوالي، فمثلاً X_{11} ترمز إلى المشاهدة الأولى في المتغيّر الأول، في حين ترمز X_{32} إلى المشاهدة الثالثة في المتغيّر الثاني و هكذا....

دعنا نحصل على $X'X$:

$$\begin{array}{ccc} X' & X & (X'X) \\ \begin{bmatrix} 1 & 1 & 1 & 1 \\ X_{11} & X_{21} & X_{31} & X_{41} \\ X_{12} & X_{22} & X_{32} & X_{42} \end{bmatrix} & \begin{bmatrix} 1 & X_{11} & X_{12} \\ 1 & X_{21} & X_{22} \\ 1 & X_{31} & X_{32} \\ 1 & X_{41} & X_{42} \end{bmatrix} = & \begin{bmatrix} n & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1^2 & \sum X_1X_2 \\ \sum X_2 & \sum X_1X_2 & \sum X_2^2 \end{bmatrix} \end{array}$$

علماً أنّ إضافة المتغيّر الوهمي (Dummy variable – Variable indicatrice)، والذي يأخذ القيم واحد (A unit vector – Vecteur d'unité) للمصفوفة X هو ضروري للحصول على قيمة الثابت b_0 [24]

دعنا الآن نحصل على ***X'Y*** :

[24] Ward, J.H., and Jennings, E., "introduction to linear Models". Prentice – Hall international, inc., 1973 pp: 26-28

$$\underset{X'}{\begin{bmatrix} 1 & 1 & 1 & 1 \\ X_{11} & X_{21} & X_{31} & X_{41} \\ X_{12} & X_{22} & X_{32} & X_{42} \end{bmatrix}} \underset{Y}{\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix}} = \underset{(X'Y)}{\begin{bmatrix} \sum Y \\ \sum X_1 Y \\ \sum X_2 Y \end{bmatrix}}$$

إذن نخلص إلى أنّه يمكننا كتابة المعادلات الطبيعية (2) و (3) و (4)، في شكل مصفوفات كالآتي:

$$(X'X)b = (X'Y)$$

وبضرب طرفي المعادلة قبلياً (premultiply - Prémultiplier) بالمصفوفة $(X'X)^{-1}$ نحصل على:

$$b = (X'X)^{-1}(X'Y)$$

وهي المعادلة المطلوبة للحصول على معاملات الانحدار الجزئية باستخدام المصفوفات للوحدات الخام.

أمّا بالنسبة لتباين توزيع المعاينة لمعامل الانحدار الجزئي، في حالة استخدام المصفوفات، فيساوي إلى:

$$S^2_{b_0} = C_{00} \cdot \frac{SS_{Resd}}{n-k-1}$$

$$S^2_{b_1} = C_{11} \cdot \frac{SS_{Resd}}{n-k-1}$$

$$S^2_{b_2} = C_{22} \cdot \frac{SS_{Resd}}{n-k-1}$$

علماً أنّ القيم C_{22}،C_{11} ، C_{00} هي القيم القطرية (diagonal values - Valeurs sur la diagonale) في مقلوب (Inverse - Inverse) المصفوفة $(X'X)$ [25]:

---

[25] Yamane, pp: 953-957

$$(X'X)^{-1} = \begin{bmatrix} C_{00} & C_{01} & C_{02} \\ C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \end{bmatrix}$$

الجدير بالذكر، أنّه في حالة استخدام المصفوفات للوحدات الخام فإنّ $\sum e^2 = e'e$، كما وأن مجموع الاختلافات الكلية للمتغيّر التابع يساوي إلى:

$$SS_T = (Y'Y) - n\overline{Y}$$

ثالثاً: الحصول على معاملات الانحدار الجزئية باستخدام مصفوفة معاملات الارتباط بين المتغيرات:

تعتبر طريقة استخدام مصفوفة معاملات الارتباط بين المتغيّرات، من أسهل، وأفضل، الطرق إطلاقاً، في الحصول على معاملات الانحدار الجزئية.

ويتم الحصول على معاملات الانحدار الجزئية بالوحدات المعيارية (Standard partial regression coefficients – Coefficients standardisés de régression partielle) باستخدام المعادلة الآتية[26]:

$$\beta = R^{-1} . V$$

علماً أنّ:

β : هي الموجه (vector - Vecteur) الذي يحتوي على معاملات الانحدار الجزئية بالوحدات المعيارية (Beta weights – Pondérations Beta).

R^{-1} : هي مقلوب (inverse - Inverse) مصفوفة معاملات الارتباط بين المتغيرات المستقلة.

V : وهي الموجه (vector - Vecteur) الذي يحتوي على معاملات الارتباط بين المتغيّر التابع وبين المتغيّرات المستقلة.

الجدير بالذكر أنّه باستطاعة الباحث تحويل معادلات الانحدار الجزئية بالوحدات المعيارية، إلى معاملات انحدار جزئية بالوحدات الخام، وذلك باستخدام المعادلة التي مرت معنا في الفصل الثاني:

$$b = \beta \frac{S_Y}{S_X}$$

[26] لاحظ أنه فيما يتعلق بتفسير (interpreting - Interprétion) معاملات الانحدار الجزئية بالوحدات المعيارية، فهو نفس التفسير الذي مر معنا عن معاملات الانحدار بالوحدات الخام، آخذين في الاعتبار أن وحدة القياس للمتغيرات هي الانحراف المعياري.

أمّا فيما يتعلّق بتباين توزيع المعاينة لمعامل الانحدار الجزئي بالوحدات الخام، فيمكننا الحصول عليه كالآتي[27]:

$$S_b^2 = \frac{\frac{SS_{\text{Re}\,sd}}{n-k-1}}{\sum X^2(1-R_j^2)}$$

علماً أنّ R_j^2 ، هي معامل التحديد المتعدد، بين المتغيّر المستقل الذي يراد تقييم أهميته النسبية، وبين بقية المتغيرات المستقلة، ويتم الحصول عليه بسهولة باستخدام المعادلة الآتية:

$$R_j^2 = 1 - \frac{1}{r^j}$$

حيث أنّ r^j هي القيمة القطرية في مقلوب مصفوفة معاملات الارتباط بين المتغيرات المستقلة R^{-1} ، والتي استخدمت في الحصول على معاملات الانحدار الجزئية بالوحدات المعيارية.

رابعاً: الحصول على معاملات الانحدار بالوحدات الخام باستخدام البواقي:

تعتبر طريقة استخدام البواقي (Residuals - Résidus) من أصعب الطرق من حيث العمليات الحسابية، لكن هذه الطريقة هي أكثر الطرق فائدة، لأنّها توضّح بشكل مفصل طبيعة ما يجري بين المتغيّرات في تحليل الانحدار المتعدد. فكما ذكرنا سابقاً، فإنّ معاملات الانحدار b_1 و b_2 تعرف باسم معاملات الانحدار الجزئية في معادلة الانحدار:

$$Y = b_0 + b_1X_1 + b_2X_2$$

حيث يقيس معامل الانحدار الجزئي b_i ، معدل التغير في Y والناتج عن تغيّر في المتغيّر المستقل بوحدة قياس واحدة مع بقاء أثر بقية المتغيّرات المستقلة الأخرى ثابتاً من الناحية الإحصائية. ويقصد بذلك، أنه لو استبعدنا أثر X_2 من X_1، ثمّ أوجدنا معامل انحدار المتغيّر التابع Y على ما تبقى من X_1 بعد حذف تأثير X_2، لحصلنا على معامل الانحدار الجزئي b_1. كذلك فلو استبعدنا تأثير X_1 من X_2، ثمّ أوجدنا معامل انحدار Y على ما تبقى من X_2 بدون تأثير X_1، لحصلنا على معامل الانحدار الجزئي b_2.

27 Kerlinger and pedhazur, pp: 65 -67

The coefficient of multiple determination –) معامل التجديد المتعدد: (Coefficient de determination multiple

يعرف معامل التحديد R^2 على أنّه نسبة التباين في الاختلافات الكلية في المتغيّر التابع، والتي تمّ تحديدها (تفسيرها) بانحدار Y على المتغيّرات المستقلة $X'S$. علماً أنّه يوجد العديد من الطرق لاحتساب قيمة R^2 وأهم هذه الطرق هي:

$$R^2 = \frac{SS_{\text{Re}\,g}}{SS_T} = 1 - \frac{SS_{\text{Re}\,sd}}{SS_T} \qquad (1)$$

دعنا نفترض أنّنا ندرس الحالة البسيطة لانحدار Y على X_1 و X_2، وعلى افتراض أنّ المتغيّرات مقاسة في شكل انحرافات عن الوسط الحسابي (للتبسيط)، فإنّ:

$$SS_{\text{Re}\,sd} = \sum e^2 = \sum e(y - \hat{y}) = \sum e(y - b_1 x_1 - b_2 x_2)$$

$$SS_{\text{Re}\,sd} = \sum ey - b_1 \sum ex_1 - b_2 \sum ex_2 \qquad (2)$$

وهنا نلاحظ أنّ $\sum ex_1$ و $\sum ex_2$ تساوي الصفر، وذلك واضح من عملية الحصول على المعادلات الطبيعية لمستوى الانحدار لأنّ:

$$\frac{\partial e^2}{\partial b_1} = -2\sum(y - b_1 x_1 - b_2 x_2)(x_1) = 0$$

إذن :

$$-2\sum ex_1 = 0$$

كذلك فإنّ:

$$\frac{\partial e^2}{\partial b_2} = -2\sum(y - b_1 x_1 - b_2 x_2)(x_2) = 0$$

إذن:

$$-2\sum ex_2 = 0$$

وبذلك تصبح المعادلة (2) كالآتي:

$$SS_{\text{Re}\,sd} = \sum e\, y \qquad (3)$$

وباستبدال e في (3) بقيمتها من معادلة الانحدار نحصل على:

$$SS_{\text{Re}\,sd} = \sum y(y - b_1 x_1 - b_2 x_2)$$

$$SS_{\text{Re}\,sd} = \sum y^2 - b_1 \sum x_1 y - b_2 \sum x_2 y \qquad (4)$$

وباستبدال قيمة $SS_{\text{Re}\,sd}$ في (1) بقيمتها من (4) نحصل على:

$$R^2 = 1 - \frac{\sum y^2 - b_1 \sum x_1 y - b_2 \sum x_2 y}{\sum y^2}$$

إذن:

$$R^2 = \frac{b_1 \sum x_1 y + b_2 \sum x_2 y}{\sum y^2} \qquad (5)$$

يتضح من المعادلة الأخيرة، إنّ إضافة أي متغيّر مستقل إلى معادلة الانحدار، سترفع من قيمة معامل التحديد المتعدد R^2، وذلك لأنّ المقام $\sum y^2$ (Denominator - Dénominateur) ثابت القيمة، مهما كان عدد المتغيّرات المستقلة. لكن تزيد قيمة البسط (Numerator - Numérateur)، بمقدار $\sum xy$ عند إضافة المتغيّر x إلى المعادلة.

وتجدر الإشارة أخيراً إلى وجود طرق أخرى للحصول على معامل التحديد المتعدد، ففي حالة استخدام المصفوفات للوحدات الخام فإنّه يمكن الحصول على R^2 كالآتي[28]:

$$R^2 = \frac{b'(X'Y) - n\bar{Y}^2}{(Y'Y) - n\bar{Y}^2}$$

علماً أنّ:

$$(Y'Y) = \sum Y^2$$

كذلك يمكن الحصول على R^2 في حالة استخدام مصفوفة معاملات الارتباط بين المتغيّرات كالآتي[29]:

$$R^2 = V_1\beta_1 + V_2\beta_2 + \ldots\ldots + V_k\beta_k$$

لكن الجدير بالذكر، هو أن إضافة أي متغيّر مستقل إلى معادلة الانحدار سترفع من قيمة معامل التحديد المتعدد، لذلك يتوجب على الباحث (ولكي يتمكن من أن يأخذ في الاعتبار الانخفاض الناتج في درجات الحرية $n-k-1$، بسبب إضافة أي متغيّر مستقل إلى معادلة الانحدار والذي من شأنه أن يجعل قيمة R^2 متحيّزة نحو الأعلى (upward biased – Biaisée vers le haut) أن يحصل على معامل التحديد المعدّل (Adjusted R^2 -R^2 ajustée) وذلك باستخدام المعادلة الآتية[30]:

$$\bar{R}^2 = 1 - \left[\left(1 - R^2\right)\frac{n-1}{n-k-1}\right] \qquad (6)$$

---

[28] Johnstom p:129

[29] Kerlinger and pedhazur, p:75

[30] المقصود بالتحيز نحو الأعلى هو أن قيمة R^2 التي نحصل عليها من بيانات العينة تعطي تقديراً أكبر من قيمة R^2 في المجتمع الإحصائي. للتوسع انظر: Kerlinger and pedhazur, pp: 282-283pp

يتضح في المعادلة (6)، أنّه عندما يزيد عدد المتغيّرات المستقلة في معادلة الانحدار، إلى أن تصبح $k = n - 1$، فإنّ قيمة $n - k - 1$ حينئذٍ تصبح صفراً، وبالتالي فلا يوجد عندئذٍ درجات حرية للاختلافات غير المفسّرة، فتصبح الاختلافات غير المفسّرة $(1 - R^2)$ مساوية للصفر، وتصبح قيمة $R^2 = 1$. ونخلص بذلك إلى أنّه إذا كان لدينا 10 مشاهدات (observations - Observations) وأخذنا 9 متغيّرات مستقلة في معادلة الانحدار فإنّ درجات الحرية $(N - k - 1) = n - (n - 1) - 1 = 0$، وبالتالي فإنّ قيمة $R^2 = 1$. كذلك فإنّه إذا كانت n صغيرة و k كبيرة مقارنة بحجم العينة n ، فإنّ $\overline{R}^2$ ستكون أصغر بكثير من R^2، حتى وأنّ قيمة $\overline{R}^2$ في هذه الحالة قد تكون سلبية (negative - négative)، بالرغم من أنّ قيمة $0 \geq R^2 \geq 1$. فلو فرضنا أنّ $n = 10$ ، $k = 6$، و $R^2 = 0.50$ ، فإنّ:

$$\overline{R}^2 = 1 - \left[(1 - 0.50) \frac{10 - 1}{10 - 6 - 1} \right] = 1 - (1.50) = -0.50$$

لذلك نجد أنّ الإحصائيين، يدعون إلى أخذ الحيطة والحذر عند استخدام الانحدار المتعدد، فيرى بعضهم أن نأخذ في تحليل الانحدار 30 مشاهدة لكل متغير مستقل في معادلة الانحدار. في حين يرى بعضهم الآخر، أن لا يقل حجم العينة عن 200 مشاهدة، وأن لا يزيد حجم العيّنة عن 600 مشاهدة. دعنا نفترض أنّ قيمة R^2 لانحدار Y على ثلاثة متغيرات مستقلة تساوي 0.36، ولنفترض أنّنا نرغب في الحصول على قيمة $\overline{R}^2$ عندما يكون حجم العيّنة مساوياً ل 10 مشاهدات، 90 مشاهدة، و150 مشاهدة على التوالي فإنّ:

$$\overline{R}^2 = 1 - \left[(1 - 0.36) \tfrac{10-1}{10-3-1} \right] = 0.19$$

$$\overline{R}^2 = 1 - \left[(1 - 0.36) \tfrac{90-1}{90-3-1} \right] = 0.34$$

$$\overline{R}^2 = 1 - \left[(1 - 0.36) \tfrac{150-1}{150-3-1} \right] = 0.35$$

أي أنّه كلما ازداد حجم العيّنة كلما اقتربت $\overline{R}^2$ من R^2. علماً أنّ $\overline{R}^2$ تستخدم كتقدير لمعامل التحديد المتعدد في المجتمع الإحصائي.

تقييم الأهمية النسبية للمتغيرات المستقلة في تحديد يباين المتغيّر التابع:

لا شك أنّ الهدف من تحليل الانحدار المتعدد هو تقييم الأهمية النسبية للمتغيّرات المستقلة في تحديد الاختلافات الكلية في المتغيّر التابع (Evaluating the relative importance of the independent variables in explaining the variation of the dependent variable – Evaluation de l'importance relative des variables indépendante dans l'explication de la variabilité de la variable dépendante). علماً أنّه باستطاعة الباحث، تقييم الأهمية النسبية للمتغيّر المستقل، وذلك باختبار جوهرية (Significance - Signification) معامل الانحدار من الناحية الإحصائية. لكن نظراً لوجود نماذج متكافئة (Equivalent models – Modèles équivalents)، ناتجة عن متغيّر مستقل دالة في متغيّر مستقل آخر(linear Dependency – Dépendance linéaire)، ونظراً لأنّ النماذج المتكافئة

b_2 b_1

أنّ $X_1 = F(x_2)$

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3$$

فانّ معامل التحديد المتعدد لهذا النموذج يُعرف باسم معامل التحديد المتعدد للنموذج التام. وباستطاعة الباحث تقييم الأهمية النسبية لأي متغيّر مستقل، وذلك بحذفه من معادلة الانحدار للنموذج التام، ثم تقييم الانخفاض في قيمة معامل التحديد المتعدد.

فإذا رغب الباحث في تقييم الأهمية النسبية للمتغيّر X_3 ، فإنّه يحذفه من النموذج التام، ويحصل على النموذج المقيّد.

$$Y = b_0 + b_1X_1 + b_2X_2$$

ويُعرف معامل التحديد لهذا النموذج باسم معامل التحديد للنموذج المقيّد. ولتقييم الانخفاض في قيمة معامل التحديد المتعدد، من النموذج التام R_F^2 إلى النموذج المقيّد R_R^2 ، فإنّه باستطاعة الباحث استخدام اختبار F – الإحصائي:

$$F = \frac{\left(R_F^2 - R_R^2\right)/\left(K_1 - K_2\right)}{\left(1 - R_F^2\right)/\left(n - k_1 - 1\right)}$$

.

علماً أنّ k_1 تمثل درجات الحرية في النموذج التام، بينما تمثل K_2 درجات الحرية في النموذج المقيّد.

ولتقييم الأهمية النسبية للمتغيّر X_2 ، فعلى الباحث أن يعيد X_3 إلى النموذج، ثم يحذف X_2 من النموذج التام ليحصل على النموذج المقيّد:

$$Y = b_0 + b_1X_1 + b_3X_3$$

ومن ثم يقيّم الانخفاض في قيمة معامل التحديد المتعدد، من النموذج التام إلى النموذج المقيّد.آخذين في الاعتبار أنّ قيمة معامل الانحدار b_1 في النموذج التام، ستختلف بالطبع عن قيمة b_1 في النموذج المقيّد. كذلك فباستطاعة الباحث أن يقيّم الأهمية النسبية لمجموعة من المتغيّرات المستقلة في آن واحد، وذلك بحذفها من النموذج التام، واختبار الانخفاض في معامل التحديد المتعدد، من النموذج التام إلى النموذج المقيّد، وذلك باستخدام اختبار F – الإحصائي. آخذين في الاعتبار أن ضرورة الحصول على النموذج المقيد لتقييم الأهمية النسبية للمتغير المستقل، تنتفي إذا لم يكن معامل التحديد للنموذج التام جوهري (Significant - Significative) من الناحية الإحصائية. علماً أن قيمة الانخفاض في معامل التحديد من النموذج التام إلى النموذج المقيد $\left(R_F^2 - R_R^2\right)$ تساوي إلى القيمة المربعة لمعامل الارتباط نصف الجزئي (semi partial correlation – semi corrélation partielle) والذي سيتم مناقشته في الفقرة القادمة[31].

[31] يدعى الأسلوب الموضّح أعلاه لتقييم الأهمية النسبية للمتغيّرات المستقلة بالحل التراجعي (Backward solution - Solution dégressive). علماً أنّه يوجد عدة طرق لهذا الغرض أهمها الحل التقدمي على خطوات (Stepwise solution – Solution pas à pas)

تجدر الإشارة أخيراً، إلى أننا سوف نحصل على نفس القيمة لمعامل التحديد R^2، مهما اختلفت تراتيب المتغيّرات المستقلة. لكن الأهمية النسبية للمتغيّرات المستقلة، سوف تختلف باختلاف تراتيبها في معادلة الانحدار، فعلى افتراض أننا ندرس انحدار Y على 10 متغيّرات مستقلة، فإنّ المتغيّر X_6 مثلاً، سيحدد جزءاً أكبر من الاختلافات الكلية في Y، إذا تمّ إدخاله في معادلة الانحدار قبل غيره من المتغيّرات، وسيحدد جزءاً أصغر في الاختلافات الكلية في Y، إذا تمّ إدخاله آخراً (بعد إدخال بقية المتغيّرات) في معادلة الانحدار، وذلك بالرغم من أنّ قيمة R^2 للنموذج التام، لن تتغيّر سواءً أدخلنا X_6 في بداية أو نهاية معادلة الانحدار. لذلك يتوجب على الباحث الاعتماد بشكل أساسي على النظرية الاقتصادية، ونتائج البحوث السابقة، وعلى الأسس المنطقية، في ترتيب متغيّراته المستقلة في معادلة الانحدار. علماً أنّ طريقة الحل التقدّمي على خطوات (Stepwise solution – Solution pas à pas) تساعد كثيراً في هذا المجال كما وأنّها تستخدم في برامج الكومبيوتر الجاهزة والمعرفة باسم SPSS (Statistical Packages of the Social Science)، والبرنامج BMD (Biomedical Statistical Package).

يتقدم الدكتور شربجي بالشكر من الدكتوراة وداد سعد لتزويدها المصطلحات باللغة الفرنسية.