الإنحدار الخطي البسيط

Simple Linear Regression

بعد الانتهاء من قراءة هذا الفصل سيكون لدى القارئ القدرة على بناء نموذج الانحدار الخطي البسيط وذلك من خلال دراسة الموضوعات التالية:

- الاختبارات الإحصائية وتشمل:
  - الاختبارات المعنوية لمعالم الانحدار الخطي البسيط.
  - فترات الثقة لمعالم الانحدار الخطي البسيط.
  - العلاقة بين فترات الثقة واختبار الفرضيات من طرفين.
- اختبار جودة الملاءمة الكلية لنموذج الانحدار الخطي البسيط ويشتمل على:
  - معامل التحديد.
  - اختبار جودة المعنوية الكلية.

1.6 مقدمة

يعتبر النموذج الخطي لمتغيرين هو الأبسط بين نماذج الانحدار المختلفة، وفي هذه الحالة يكون اهتمامنا مركزاً على وصف العلاقة الخطية التي تربط بين متغيرين فقط، أحدهما تابع، والآخر مستقل. وبصورة عامة إذا رمزنا للمتغير التابع بالرمز (Y) وللمتغير المستقل بالرمز (X) فإن نموذج الانحدار الخطي البسيط يكون على النحو التالي:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i , \qquad for\ i = 1,2,\ldots,n \qquad (6.1)$$

حيث: β_0 ، β_1 معالم مجهولة القيم وثوابت تختص بالمجتمع.

β_0: الجزء المقطوع من محور Y الرأسي ويسمى الحد الثابت للنموذج.

β_1: ميل الدالة الخطية ويسمى الميل الحدي للنموذج.

ε: حد الخطأ (العنصر) العشوائي.

n: عدد المشاهدات.

أسباب وجود حد الخطأ العشوائي ε:

- وجود عدة متغيرات مستقلة لها تأثير معين على المتغير التابع Y، وقد تم استبعادها من العلاقة الخطية في المعادلة (6.1) وتم احتوائها في المتغير العشوائي ε.
- وجود أخطاء ممكنة في قياس المتغير التابع Y تم احتواء تأثيرها في المتغير العشوائي ε.
- وجود خطأ تجريبي نتيجة للتجربة أو القياس من قبل الباحث تم احتواء تأثيره في المتغير العشوائي ε.

2.6 الاختبارات الإحصائية

1.2.6 الاختبارات المعنوية لمعالم الانحدار الخطي البسيط

بفرض أنه لدينا نموذج الانحدار الخطي البسيط في معادلة (6.1) لاختبار الفرضية الصفرية $\mathrm{H}_0 : \beta_i = \beta_{\mathrm{H}_0}$ مقابل الفرضية البديلة:

- $\mathrm{H}_1 : \beta_i \neq \beta_{\mathrm{H}_0}$
- $\mathrm{H}_1 : \beta_i > \beta_{\mathrm{H}_0}$
- $\mathrm{H}_1 : \beta_i < \beta_{\mathrm{H}_0}$

فإننا نستعمل إحصاء الاختبار:

$$\mathrm{T}_i = \frac{\hat{\beta}_i - \beta_{\mathrm{H}_0}}{SE(\hat{\beta}_i)} \ , \ i = 0,1 \tag{6.2}$$

حيث أن:

$\hat{\beta}_0$: القيمة المقدرة للجزء المقطوع من محور Y (الثابت).

$\hat{\beta}_1$: قيمة معامل الانحدار المقدرة للمتغير المستقل.

β_{H_0} : قيمة β_i بفرض أن H_0 صحيحة.

$SE\left(\hat{\beta}_i\right)$: الخطأ المعياري لقيمة معامل الانحدار المقدرة $\hat{\beta}_i$.

مع العلم بأن إحصاء الاختبار في (6.2) يخضع لتوزيع T بدرجات حرية $(n-2)$.

حالة خاصة: إذا كانت $\beta_{\mathrm{H}_0} = 0$ ، فإن إحصاء الاختبار يصبح على النحو التالي:

$$\mathrm{T}_i = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \ , \ i = 0,1 \tag{6.3}$$

وكذلك إحصاء الاختبار في (6.3) يخضع لتوزيع T بدرجات حرية $(n-2)$.

2.2.6 فترات الثقة لمعالم الانحدار الخطي البسيط

تعتبر فترة الثقة من الأدوات القوية التي تعطي معلومات عن المعلمة المجهولة مثلاً (β_i) باستعمال العينة. فترة الثقة نهايتاها متغيران عشوائيان، أي أنها فترة عشوائية تحاول أن تحتوى المعلمة المجهولة β_i. مع العلم أن فترة الثقة تفسَّر على أنها التكرار النسبي لمحاولات المعاينة الكبيرة والمتكررة. بفرض أن 95% مثلاً من فترات الثقة ستحتوى على β_i وأن 5% لا تحتويها، وبالتالي فإن تفسير فترة الثقة 95% للمعلمة β_i يعني أنه إذا أخذت مائة عينة عشوائية حجمها n وفي كل مرة نحسب $\hat{\beta}_i$ ونحسب فترة الثقة فإننا نتوقع أن 95 فترة تحوى على قيمة β_i الحقيقة.

فترة الثقة $100(1-\alpha)\%$ **للمعلمة** β_i **هي:**

$$\hat{\beta}_i - t(1-\frac{\alpha}{2};n-2)SE(\hat{\beta}_i), \hat{\beta}_i + t(1-\frac{\alpha}{2};n-2)SE(\hat{\beta}_i) \qquad (6.4)$$

مع ملاحظة أن $t(1-\frac{\alpha}{2};n-2)$ يمكن حسابها من خلال جداول خاصة بتوزيع T .

3.2.6 العلاقة بين فترات الثقة واختبار الفرضيات من طرفين

إذا احتوت فترة الثقة على القيمة الفرضية (β_{H_0}) فإننا لا نرفض الفرضية الصفرية والعكس صحيح.

3.6 اختبار جودة الملاءمة الكلية لنموذج الانحدار الخطي البسيط

كلما كانت المشاهدات أقرب إلى خط الانحدار زاد الاختلاف (التباين) الكلي في المتغير التابع والذي تفسره معادلة خط الانحدار. ويمكن قياس جودة ملائمة (توفيق) ومعنوية نموذج الانحدار بطريقتين هما معامل التحديد واختبار تحليل التباين. وسنتناول كل منهما فيما يلي:

1.3.6 معامل التحديد Coefficient of Determination

معامل التحديد يمثل النسبة بين مجموع مربعات الانحدار ومجمـوع المربعـات الكلـي، ويرمز له بالرمز R^2 ويمثل نسبة التغير الكلي في المتغير التابع والتي يمكن تفسيرها مـن خلال نموذج الانحدار المُقدر، وإشارته دائما موجبة محـصورة بـين الـصفر والواحـد الصحيح، أي أن:

$$0 \leq R^2 \leq 1$$

ملاحظات على معامل التحديد

- إذا كانت قيمة معامل التحديد تساوي الصفر، عندها لا يُفسر نموذج الانحدار شـيئاً من التباين في المتغير التابع.
- إذا كانت القيمة المُطلقة لمعامل التحديد تساوي الواحد الصحيح، عندها تقع كل نقاط الانتشار على خط الانحدار المقدّر، وهذا نادر الحدوث في التطبيقات العملية.
- غالباً قيمة معامل التحديد تزيد عن الصفر وتقل عن الواحد الصحيح.
- معامل التحديد يساوى مربع معامل الارتباط الخطي لبيرسون في حالـة الانحـدار الخطي البسيط.
- حيث أن مقدرات طريقة المربعات الصغرى العادية يمكن الحصول عليها من خلال تصغير(تقليل) مجموع مربعات البواقي (الأخطاء)، فإن هذه الطريقة تعطـى أكبـر قيمة معامل لمعامل التحديد في حالة النماذج الخطية.
- يمكن الحصول على قيمة معامل بيرسون للارتباط الخطي البسيط (r) وذلك بأخـذ الجذر التربيعي لمعامل التحديد، أي أن: $r = \pm\sqrt{R^2}$.

- معامل بيرسون للارتباط الخطي البسيط يستخدم لمعرفة نوع واتجاه العلاقة الخطية بين متغيرين كميين.

ملاحظات على معامل ارتباط بيرسون الخطي

- معامل الارتباط له قيمة واحدة سواء أكان X أم Y هو المتغير التابع أي أن:

$$r_{XY} = r_{YX} \tag{6.5}$$

- العلاقة بين المتغيرين X و Y هي علاقة خطية، ويمكن الاستدلال على نوع العلاقة بين المتغيرين موضوع الدراسة من خلال مسار النقاط التـي تمثـل أزواج القـيم المتناظرة في شكل لوحة الانتشار.
- معامل الارتباط يكون مستقلاً عن وحدات القياس لكل من المتغيرين.
- يستخدم في حالة المتغيرات الكمية.
- يستخدم في قياس الارتباطات الخطية فقط، ولا يـستخدم فـي الارتباطـات غيـر الخطية.
- لا يُفسر السببية إذا يدرس فقط اتجاه ودرجة العلاقة بين المتغيرين.
- إذا كانت $r = \pm 1$ فهذا يعني أن العلاقة بين المتغيرين طردية (عكسية) تامـة، أمـا إذا كانت $r = 0$ فهذا يدل على عدم وجود علاقة خطية بين المتغيرين.

2.3.6 اختبار جودة المعنوية الكلية

يستخدم اختبار F (نسبة للعالم Fisher) لاختبار المعنوية الكلية لنموذج الانحدار الخطي البسيط، ويستخدم لاختبار الفرضية الصفرية $\mathrm{H}_0 : \beta_1 = 0$، وهذا يكافئ اختبار T السابق

•

تطبيق عملي (1.6):

البيانات التالية تختص بإجمالي الإنفاق الاستهلاكي (Y) مقاسا بمليارات الدولارات وإجمالي الدخل المتاح (X) مقاسا بمليارات الدولارات أيضاً لاقتصاد معين في الفترة 2000 – 2012. اسم الملف (Example6.1).

جدول (1.6): إجمالي الإنفاق الاستهلاكي والدخل المتاح

السنة	Y	X
2000	85	95
2001	91	108
2002	98	120
2003	103	128
2004	109	139
2005	114	145
2006	119	153
2007	122	164
2008	133	175
2009	140	180
2010	145	187
2011	163	290

المطلوب:

1. ارسم لوحة الانتشار.
2. اختر النموذج المناسب الذي يعبر عن العلاقة بين الإنفاق الاستهلاكي (Y) والدخل المتاح (X).
3. أوجد معادلة الانحدار الخاصة بذلك النموذج واكتبه بالشكل القياسي المناسب.
4. ارسم خط الانحدار.

الحل:

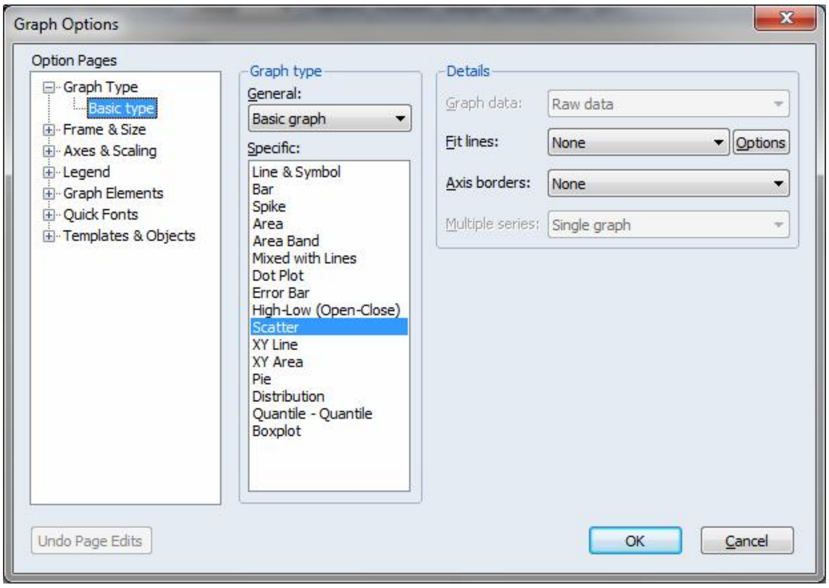
1. رسم لوحة الانتشار:

لرسم شكل الانتشار نتبع الخطوات التالية من خلال برنامج E-Views:

- أولاً: اختر المتغيرين X، Y ثم اضغط على مفتاح الإدخال، أو من خلال التالي:

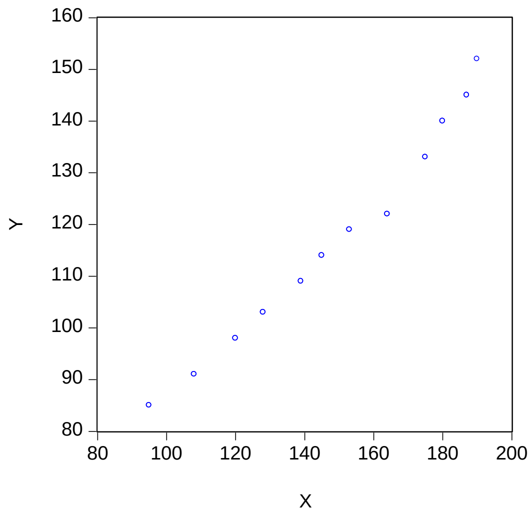
View ► Open Selected ► One Window ► Open group

- ثانياً: اختر `View` من شريط الاختبارات وذلك في نافذة عرض البيانات الخاصة بالمتغيرين X,Y ثم اختر `Graph`.
- ثالثاً: اختر `Scatter` أسفل قائمة `Graph Type` كما في شكل (1.6).



شكل (1.6): المربع الحواري للاختيار Graph

- اضغط OK، نحصل على الرسم الموضح في شكل (2.6).

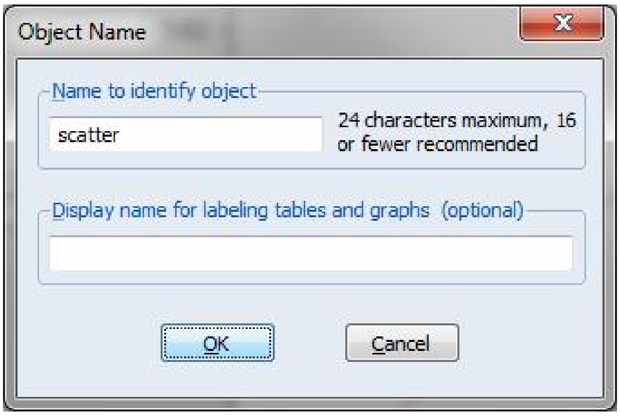


شكل (2.6): شكل الانتشار لنموذج انحدار الإنفاق الاستهلاكي والدخل

2. النموذج المناسب

يمكن ملاحظة من الشكل (2.6) أن النموذج الخطي يعتبر مناسباً في هذه الحالة حيث يتبين أن هنالك اتجاهاً خطياً عاماً واضحاً في العلاقة بين الإنفاق الاستهلاكي والدخل المتاح.

- لحفظ الرسم الحالي اختر ***Name*** من شريط الاختيارات ثم اكتب اسماً مناسباً مثلاً ***scatter*** كما هو موضح في شكل (3.6).



شكل (3.6): المربع الحواري الخاص بتسمية الرسم البياني باسم scatter

3. معادلة الانحدار

لإيجاد معادلة الانحدار الخطي نتبع الخطوات التالية في برنامج:E Views

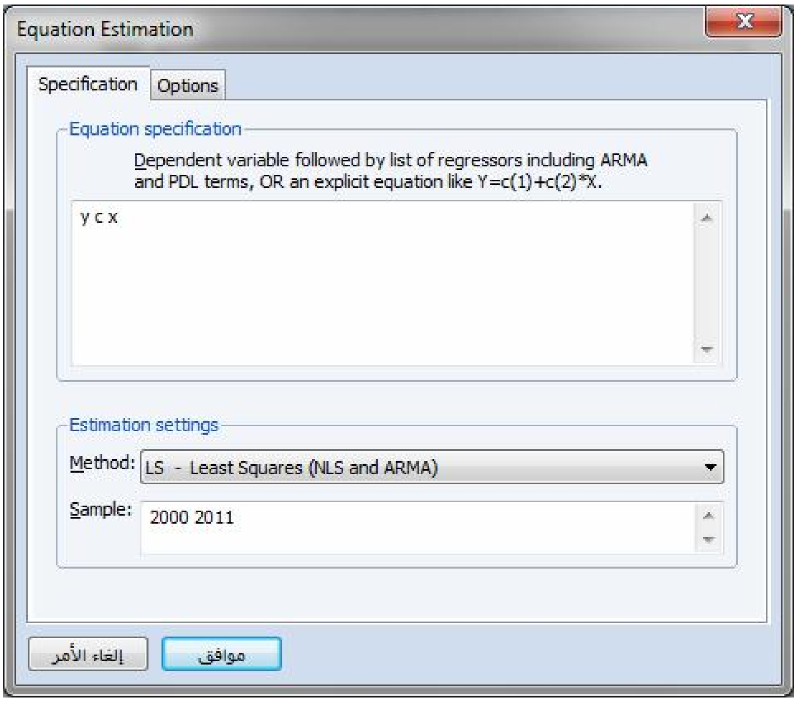
- من شريط القوائم اختر

Quick ► Estimate Equation

- ندخل معادلة خط الانحدار الخطي كما يلي:

Y C X

كما هو موضح في شكل (4.6).



شكل (4.6): المربع الحواري كتابة نموذج انحدار الإنفاق الاستهلاكي والدخل

حيث: Y هو المتغير التابع ثم يتبعه C والذي يمثل الجزء الثابت (المقطوع من محور Y) ثم المتغير المستقل X.

- يجب ملاحظة أن الترتيب ضروري في هذه الحالة حيث يجب أن نبدأ بكتابة المتغير التابع ثم C للدلالة على الجزء الثابت ثم المتغير المستقل أو المتغيرات المستقلة (كما سيأتي شرحه في الفصل السابع إن شاء الله تعالى).
- اضغط موافق سنحصل على النتائج الموضحة في جدول (2.6).
- لحفظ النتائج الحالية اختر ***Name*** من شريط الاختيارات ثم اكتب اسماً مناسباً مثلا ***EQ1*** كما تم شرحه سابقاً.
- وبذلك تكون معادلة انحدار الإنفاق الاستهلاكي المُقدّرة هي:

$$\hat{y}_i = 16.942 + 0.677 X_i \qquad (6.6)$$

جدول (2.6): نتائج نموذج انحدار الإنفاق الاستهلاكي والدخل

Dependent Variable: Y
Method: Least Squares
Date: 03/31/13 Time: 06:36
Sample: 2000 2011
Included observations: 12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	16.94154	4.784922	3.540610	0.0054
X	0.676963	0.031538	21.46472	0.0000

R-squared	0.978757	Mean dependent var	117.5833
Adjusted R-squared	0.976632	S.D. dependent var	21.63523
S.E. of regression	3.307273	Akaike info criterion	5.381137
Sum squared resid	109.3806	Schwarz criterion	5.461955
Log likelihood	-30.28682	Hannan-Quinn criter	5.351215
F-statistic	460.7342	Durbin-Watson stat	0.755642
Prob(F-statistic)	0.000000		

تفسير معاملي الانحدار في المعادلة (6.6):

- $\hat{\beta}_0 = 16.942$: قيمة الإنفاق الاستهلاكي الإجمالي يساوي 16.942 بليون دولار عندما يكون الدخل المتاح يساوي صفراً.
- $\hat{\beta}_1 = 0.677$: قيمة ميل خط الانحدار المُقدّر، ويفسر الميل الحدي للاستهلاك أو التغير في الاستهلاك الناتج من تغير وحدة واحدة من الدخل، وهذا يعني أنه إذا زاد الدخل المتاح بمقدار وحدة واحدة (أي بليون دولار) فإن الاستهلاك يزداد بمقدار 0.677 بليون دولار أي 677 مليون دولار.
- ويمكن ملاحظة أن $\hat{\beta}_0 > 0,\ \ 0 < \hat{\beta}_1 < 1$ حسب نظرية الدخل المطلق.
- كما يمكننا قياس المرونة الداخلية للاستهلاك η والتي تقيس الاستجابة النسبية في الاستهلاك للتغيرات النسبية في الدخل المتاح حسب القانون:

$$\eta_i = \frac{dY_i}{dX_i} . \frac{X_i}{Y_i} = \hat{\beta} \frac{X_i}{Y_i},\ i = 1, 2, \ldots, n \qquad (6.7)$$

حيث أن المرونة تتغير حسب نقطة العينة (i) فإننا غالباً نستخدم نقطة المتوسطات للحصول على معامل المرونة عند الوسط:

$$\eta = \frac{dY}{dX} . \frac{\overline{X}}{\overline{Y}} = \hat{\beta} \frac{\overline{X}}{\overline{Y}} \qquad (6.8)$$

حساب قيمة المتوسط الحسابي:

نختار المتغيرين X,Y ثم اضغط مفتاح الإدخال **Enter**

View ► Descriptive Statistics ► Common Sample

فنحصل على النتائج الموضحة في جدول(3.6).

جدول (3.6): نتائج الإحصاء الوصفي للمتغيرين الإنفاق الاستهلاكي والدخل

	X	Y
Mean	148.6667	117.5833
Median	149.0000	116.5000
Maximum	190.0000	152.0000
Minimum	95.00000	85.00000
. Std. Dev	31.61798	21.63523
Skewness	-0.226810	0.109924
Kurtosis	1.833353	1.856667
Jarque-Bera	0.783418	0.677772
Probability	0.675901	0.712564
Sum	1784.000	1411.000
. Sum Sq. Dev	10996.67	5148.917
Observations	12	12

من خلال جدول (3.6) تبين أن:

$$\bar{X} = 148.667 \ , \bar{Y} = 117.583$$

بالتعويض في المعادلة (6.8) نجد أن معامل المرونة عند الوسط يساوي

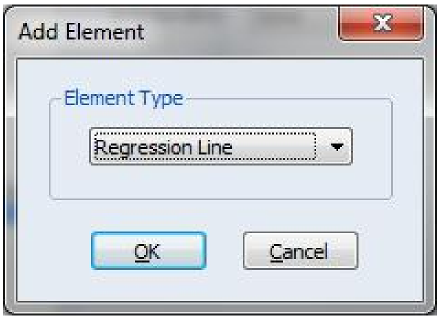
$$\eta = 0.677 \times \frac{148.667}{117.583} = 0.856$$

مع ملاحظة أن المرونة ليس لها وحدة قياس وقد تكون عكسية أو طردية حسب طبيعة العلاقة التي تربط المتغيرين.

4. رسم خط الانحدار

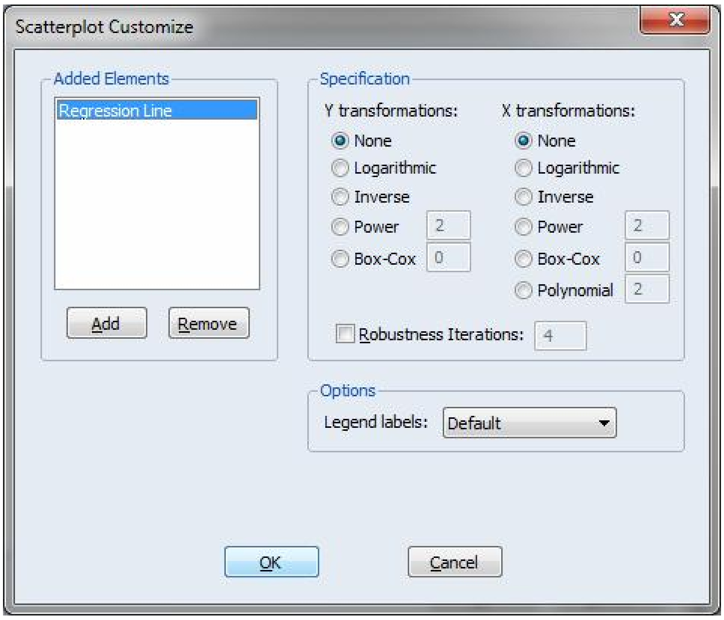
لرسم خط الانحدار نتبع الخطوات التالية:

- من شكل (1.6) اختر Options من خلال Fit lines فيظهر المربع الحواري في شكل (5.6).



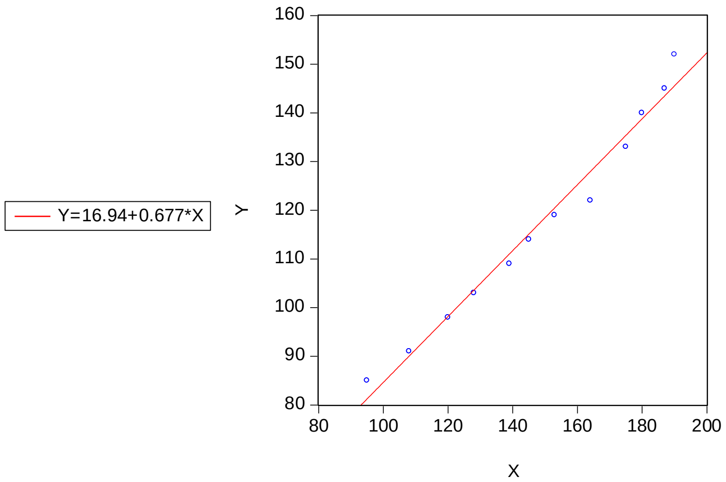
شكل (5.6): المربع الحواري الخاص بخط الانحدار

▪ اختر Regression Line أسفل Element Type ثم OK فيظهر المربع الحواري في شكل (6.6).



شكل (6.6): المربع الحواري الخاص بتخصيص خط الانحدار - 1

- اختر Detailed مقابل Legend labels
- اضغط OK فيظهر المربع الحواري في شكل (6.6) ثم اضغط OK فنحصل علـــى الرسم الموضح في شكل (7.6).



شكل (7.6): المربع الحواري الخاص بتخصيص خط الانحدار - 2

- أحفظ الرسم بنفس الاسم السابق Scatter

اختبر معنوية كل من β_1, β_0 باستخدام اختبـــار t

وذلك عند مستوى دلالة 5%.

الحل:

.

افتح الملف Example6.1 ثم اضغط ***EQ1*** فنحصل على النتائج الموضحة في جدول (2.6). من خلال نتائج جدول (2.6) نستنتج ما يلي:

- اختبار معنوية المعلمة β_0:

$$H_0: \beta_0 = 0 \qquad Versus \qquad H_1: \beta_0 \neq 0$$

قيمة t المحسوبة تساوي 3.541 وقيمة الاحتمال (Prob. Value) تساوي 0.0054، حيث أن قيمة الاحتمال أقل من مستوى الدلالة 5%؛ بالتالي نرفض الفرضية الصفرية القائلة بأن $\beta_0 = 0$ أي أن β_0 معنوية إحصائياً عند مستوى دلالة 5%.

- اختبار معنوية المعلمة β_1:

$$H_0: \beta_1 = 0 \qquad Versus \qquad H_1: \beta_1 \neq 0$$

قيمة t المحسوبة تساوي 21.465 وقيمة الاحتمال (Prob. Value) تساوي 0.0000، حيث أن قيمة الاحتمال أقل من مستوى الدلالة 5%؛ بالتالي نرفض الفرضية الصفرية القائلة بأن $\beta_1 = 0$ أي أن β_1 معنوية إحصائياً عند مستوى دلالة 5%، وهذا يعني أن متغير الدخل المتاح دال إحصائيا.

أوجد فترة الثقة 95% لكل من β_0، β_1

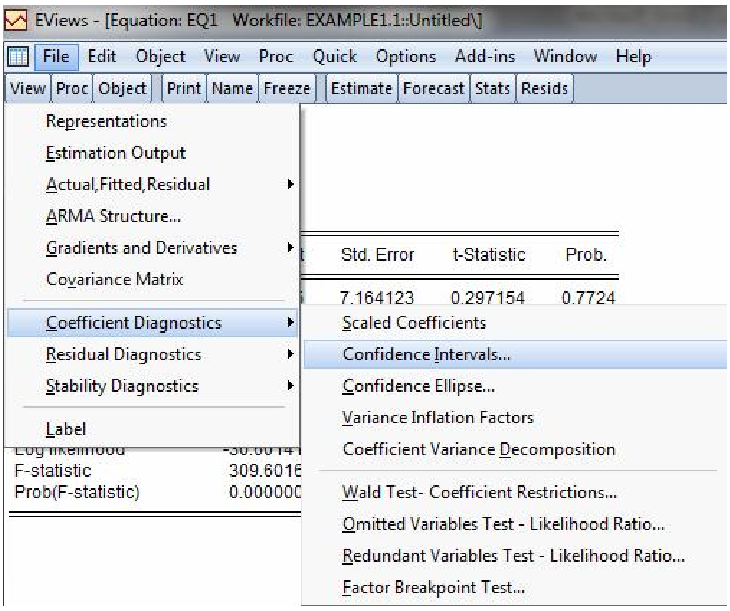
الحل:

(2.6). لإيجاد فترة الثقة نتبع التالي:

- من نافذة نتائج الانحدار اختر التالي:

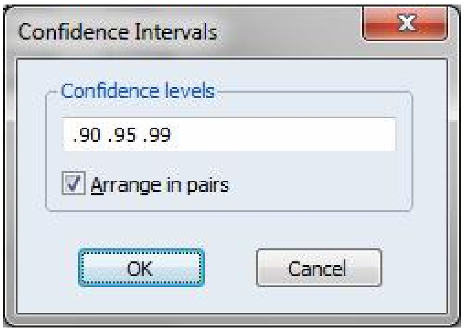
View ► Coefficient Diagnostic ►Confidence Intervals

كما يظهر في الشكل (8.6).



شكل (8.6): المربع الحواري الخاص بفترات الثقة

▪ يظهر المربع الحواري كما في شكل (9.6).



شكل (9.6): المربع الحواري الخاص بمستويات الثقة

▪ أكتب 0.95 في أسفل Confidence Intervals ثم OK فنحصل على النتائج الموضحة في جدول (4.6)

جدول (4.6): نتائج فترة الثقة لنموذج انحدار الإنفاق الاستهلاكي والدخل

Coefficient Confidence Intervals
Date: 03/31/13 Time: 06:39
Sample: 2000 2011
Included observations: 12

		95% CI	
Variable	Coefficient	Low	High
C	16.94154	6.280071	27.60301
X	0.676963	0.606691	0.747235

▪ فترة 95% ثقة للمعلمة β_0 هي (6.280 , 27.603) بمعنى أن β_0 تقع في المدى $6.280 < \beta_0 < 27.603$.

▪ يلاحظ أن فترة الثقة للمعلمة β_0 لا تشتمل على الصفر وهذا يعني رفض الفرضية الصفرية القائلة بأن $H_0: \beta_0 = 0$ وهذا يدل على أن β_0 تعتبر معنوية إحصائياً.

- فترة 95% ثقة للمعلمة β_1 هي (0.607, 0.747) بمعنى أن β_1 تقع في المـــدى $0.607 < \beta_1 < 0.747$.
- يلاحظ أن فترة الثقة للمعلمة β_1 لا تشمل على الصفر وهذا يعني رفض الفرضـــية الصفرية القائلة بأن $H_0 : \beta_1 = 0$ وهذا يدل على أن β_1 تعتبر معنوية إحصائياً.

ملاحظة: إذا اشتملت فترة الثقة على الصفر فإننا لا نرفض الفرضية الصفرية وتكون المعالم في هذه الحالة غير معنوية إحصائياً

المطلوب:

1. أوجد قيمة معامل التحديد مع تفسير النتيجة.
2. أحسب معامل الارتباط مع تفسير النتيجة.
3. هل نموذج الانحدار المقدر يصلح للتنبؤ ومناسب للبيانات؟
4. أوجد قيمة إجمالي الإنفاق الاستهلاكي (Y) المقدرة إذا علمت أن الدخل المتاح يساوي 153 بليون دولار.
5. أوجد قيمة الخطأ في التقدير فـــي البنـــد (4) علمـــاً بـــأن القيمـــة الفعليـــة للإنفـــاق الاستهلاكي 119 بليون دولار.

الحل:

بالرجوع إلى النتائج الموضحة في جدول (2.6)، يمكن استنتاج ما يلي:

1. قيمة معامل التحديد تساوي $R^2 = 0.9788$ وهذا يعني أن 97.88% من التغير أو التباين في الإنفاق الاستهلاكي تم تفسيره من خلال الدخل المتـــاح أو مـــن خـــلال

معادلة الانحدار، أما النسبة المتبقية 2.12% فترجع إلى وجود متغيـــرات مـــستقلة أخرى قد تؤثر في الإنفاق الاستهلاكي.

2. لإيجاد معامل الارتباط نستخدم المعادلة $r = \pm\sqrt{R^2}$، حيـــث أن إشـــارة ميـــل الانحدار موجبة ($\hat{\beta}_1 = 0.677$) بالتالي فإن:

$$r = \sqrt{0.9788} = 0.9893$$

وهذا يدل على وجود ارتباط طردي قوي بين الدخل المتاح والاستهلاك.

3. حيث أن قيمة $F = 460.734$ وقيمة الاحتمال (Prob. Value) تـــساوي 0.000 وهي أقل من مستوى الدلالة 95%، لذلك نرفض الفرضية الصفرية القائلـــة بـــأن النموذج المقدر لا يصلح للتنبؤ، وبالتالي يمكن القول بان نموذج الانحـــدار المقـــدر يصلح للتنبؤ. مع ملاحظة أن قيمة F تساوي مربع قيمـــة t وذلـــك فـــي حالـــة الانحدار البسيط أي أن: $F = t^2$، $F = (21.465)^2 = 460.734$

4. التقدير

بالتعويض في المعادلة (6.6) نجد أن:

$$\hat{Y} = 16.942 + 0.677(153) = 120.523$$

5. الخطأ في التقدير

$$e = 119 - 120.523 = -1.523$$