المحور الثاني: الانحدار الخطي البسيط

أولا: تقديم نموذج الانحدار الخطي البسيط

1- الشكل العام: النموذج الخطي البسيط يأخذ الشكل التالي:

$$Y_t = \alpha + \beta \cdot X_t + \varepsilon_t$$

سمي النموذج خطيا لأن العلاقة بين المتغير التابع والمستقل علاقة خطية، وسمي البسيط لأن عدد المتغيرات المستقلة متغير واحد فقط، و $\alpha \,\&\, \beta$ معلمات أو معاملات النموذج.

حيث: Y : المتغير التابع، أو المتغير الداخلي.

X : المتغير المفسر، أو المتغير المستقل.

ε : المتغير العشوائي.

α و β: معلمات للتقدير.

t : مؤشر الزمن.

تمثل المعلمة α الجزء الثابت، وهو الجزء المقطوع من المحور الرأسي، وهو عبارة عن قيمة متوسط المتغير التابع لما تنعدم قيمة المتغير المستقل، بينما تمثل المعلمة β معامل الانحدار أو ميل الخط المستقيم، وتعبر عن مقدار التغير في المتغير التابع نتيجة لتغير المتغير المستقل بوحدة واحدة، وتبين اشارتها إذا ما كانت العلاقة بين المتغير التابع والمستقل علاقة طردية أو عكسية.

إن إدخال المتغير العشوائي ε_t في النموذج القياسي له عدة مسوغات أهمها أنه عبارة عن مجموعة شاملة تتضمن كل تلك المتغيرات التي لا يمكن قياسها بسهولة، قد يمثل هذا الحد المتغيرات التي لا يمكن إدراجها في النموذج لعدم توفر البيانات، أو أخطاء في القياس في البيانات، أو العشوائية الموجودة في السلوك البشري.

2- الفرضيات الاحتمالية التي يقوم عليها النموذج الخطي البسيط:

إن الطريقة المستعملة في تقدير معاملات نموذج الانحدار الخطي البسيط هي طريقة المربعات الصغرى العادية "OLS"، التي تم وضعها من طرف : Carl Friedrich Gauss، بناءا على بعض الفرضيات التي تجعل منها الطريقة الأكثر استعمالا، و تدور هذه الفرضيات حول طبيعة وشكل المتغير العشوائي، وهي:

- $\forall t \quad E(\varepsilon_t) = 0$: وتنص هذه الفرضية على أن الأخطاء لا تدخل في تفسير Y، حيث تعبر عن قيم عشوائية تأخذ قيما سالبة، موجبة أو معدومة، لا يمكن قياسها وتحديدها بدقة، تخضع للقوانين الاحتمالية، حيث أن أملها الرياضي أو متوسطها يكون معدوما.
- $\forall t \quad V(\varepsilon_t) = E(\varepsilon_t^2) = \delta_\varepsilon^2$: ثبات أو تجانس التباين (Homoscedasticity). أي أن تشتت الأخطاء حول متوسطها معدوم .
- $\forall i \neq j \quad Cov(\varepsilon_i \varepsilon_j) = E(\varepsilon_i \varepsilon_j) = 0$: عدم وجود ارتباط ذاتي بين الأخطاء، أي أن التباينات المشتركة بين الأخطاء تكون معدومة.
- $Cov(x_t . \varepsilon_t) = 0$: عدم وجود ارتباط بين المتغير المستقل والمتغير العشوائي.
- $\varepsilon_t \to N(0 . \delta_\varepsilon^2)$: التوزيع الاحتمالي للمتغير العشوائي هو التوزيع الطبيعي.

فرضيات أخرى:

- المتغيرات Y و X محددة بدون خطأ.
- قيم المتغير X غير عشوائية.

ثانيا: مقدرات معلمات النموذج الخطي البسيط:

يتم تقدير نموذج الانحدار الخطي البسيط بطريقة المربعات الصغرى العادية (Ordinary Least Squares)، التي تهدف إلى الحصول على مقدرات $\hat{\alpha}$. $\hat{\beta}$ تعطي مجموع مربعات انحراف القيم المقدرة عن القيم الحقيقية في أدنى قيمة لها.

مقدرات المعلمات بطريقة التقدير OLS (بدون برهان رياضي) مساوية دائما لـ:

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \cdot \bar{X}$$

$$\hat{\beta} = \frac{\sum X_t Y_t - n\bar{Y}\bar{X}}{\sum X_t^2 - n\bar{X}^2} = \frac{\sum (Y_t - \bar{Y}) \cdot (X_t - \bar{X})}{\sum (X_t - \bar{X})^2} = \frac{Cov(X_t . Y_t)}{V(X_t)}$$

مثال:

ترغب إحدى الشركات في تحديد العلاقة بين إنفاقها على الدعاية والاعلانات وعوائد المبيعات، كلاهما بالمليون دينار جزائري، فإذا كانت لدينا البيانات التالية عن تطور هاذين المتغيرين من 2014 إلى 2023 كمايلي:

السنة	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
الاعلانات	4	5	6	6	7	8	7	9	8	10
المبيعات	44	42	52	48	50	60	58	62	64	70

المطلوب:

1- إدخال البيانات يدويا في برنامج EViews.

2- أنشئ ملف خارجي Excel خاص بهذه البيانات، ثم قم بتصديره إلى البرنامج الاحصائي EViews.

3- بالاستعانة ببرنامج EViews:

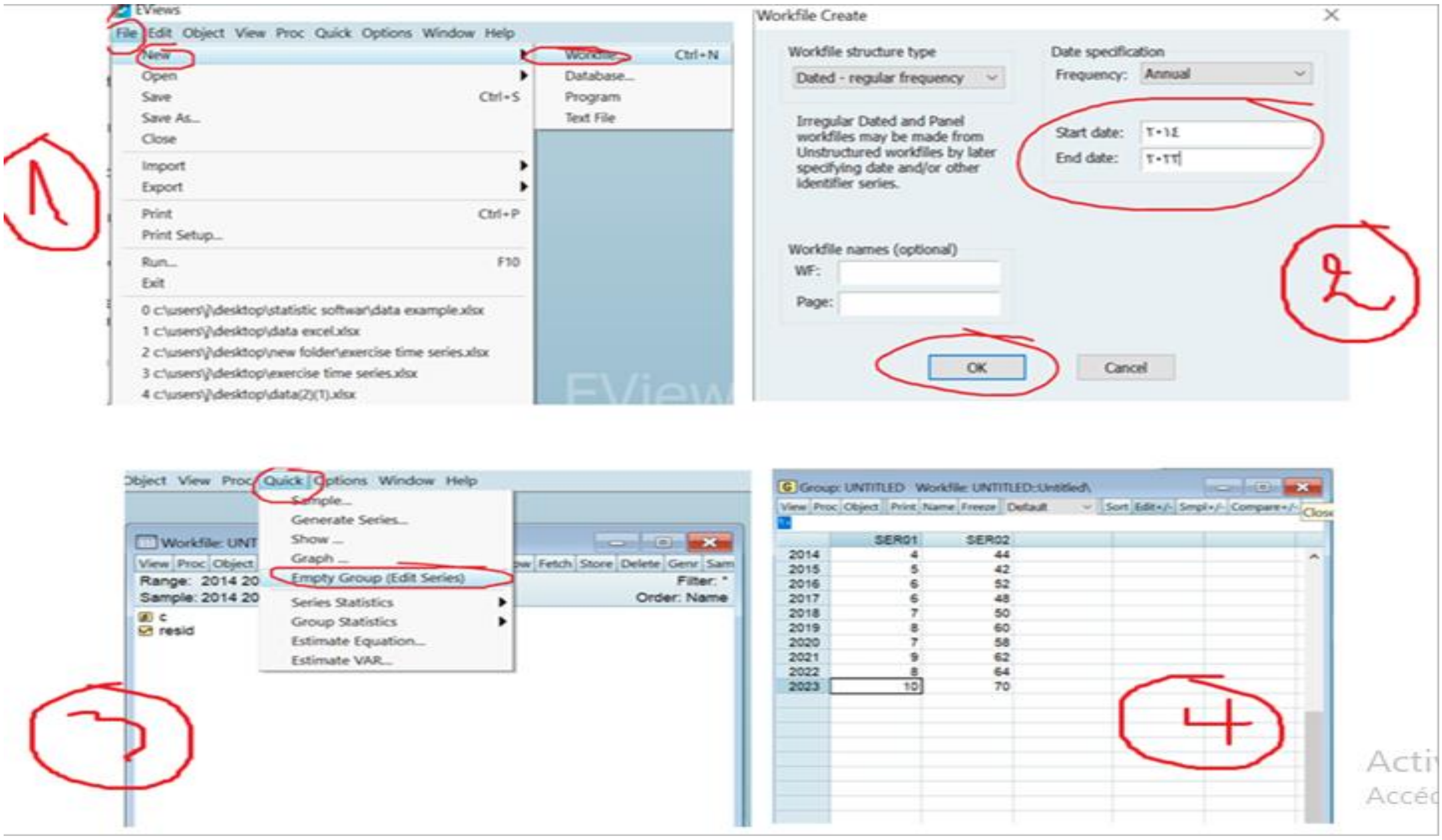
أ- مثل بيانيا بيانات الجدول بسحابة النقاط، ماذا تستنتج؟

ب- قدر النموذج الخطي البسيط الذي يقيس أثر الانفاق على الاعلانات على عوائد المبيعات في هذه الشركة، وفسر النتائج.

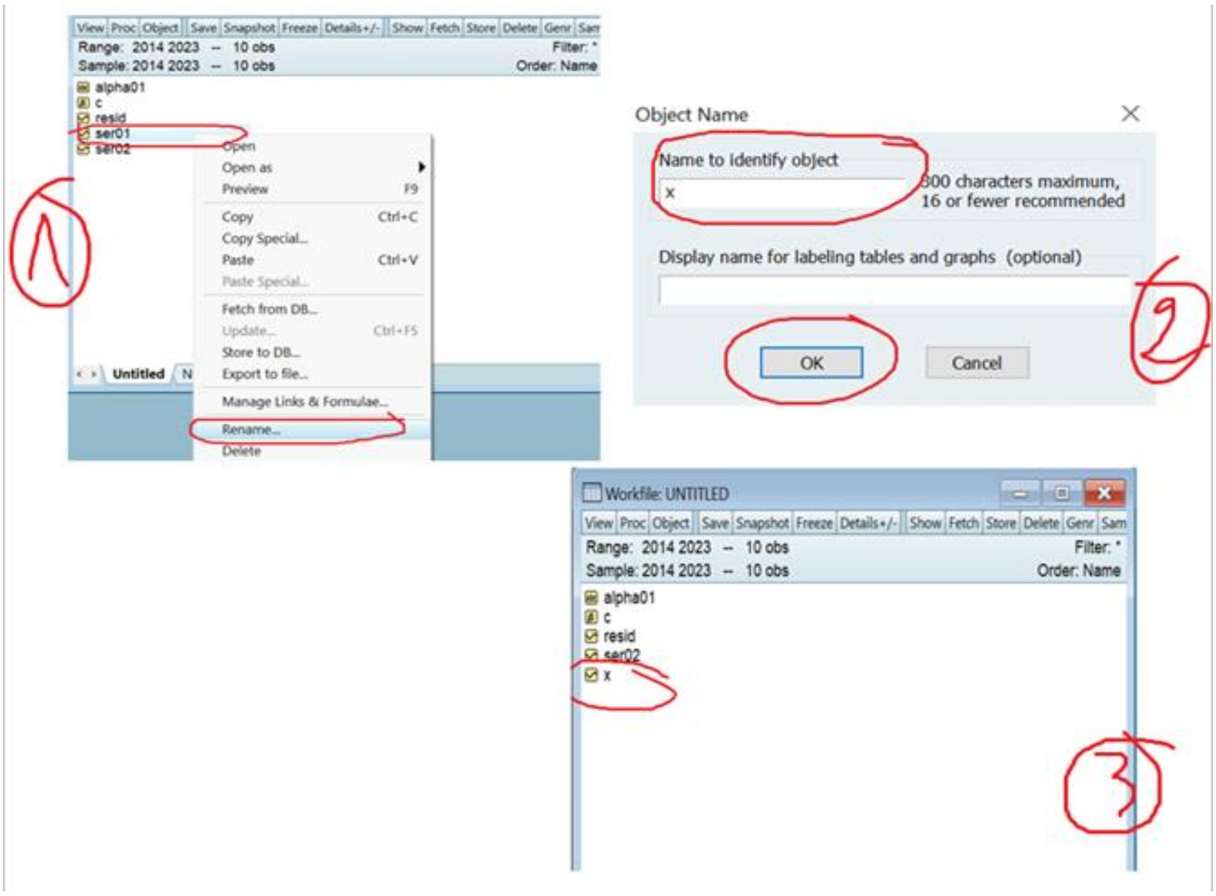
ت- حساب القيم المقدرة $\hat{Y}_t$ واستنتاج بواقي التقدير e_t.

الحل:

1- **ادخال لبيانات يدويا:** يتم من خلال الأوامر التالية المرقمة من 1-4



بعد الانتهاء من المرحلة 4 نقوم بتسمية المتغيرات وذلك بالضغط على الزر الأيمن للفأرة فيظهر مايلي:



2) تصدير البيانات من خلال ملف خارجي EXCEL:

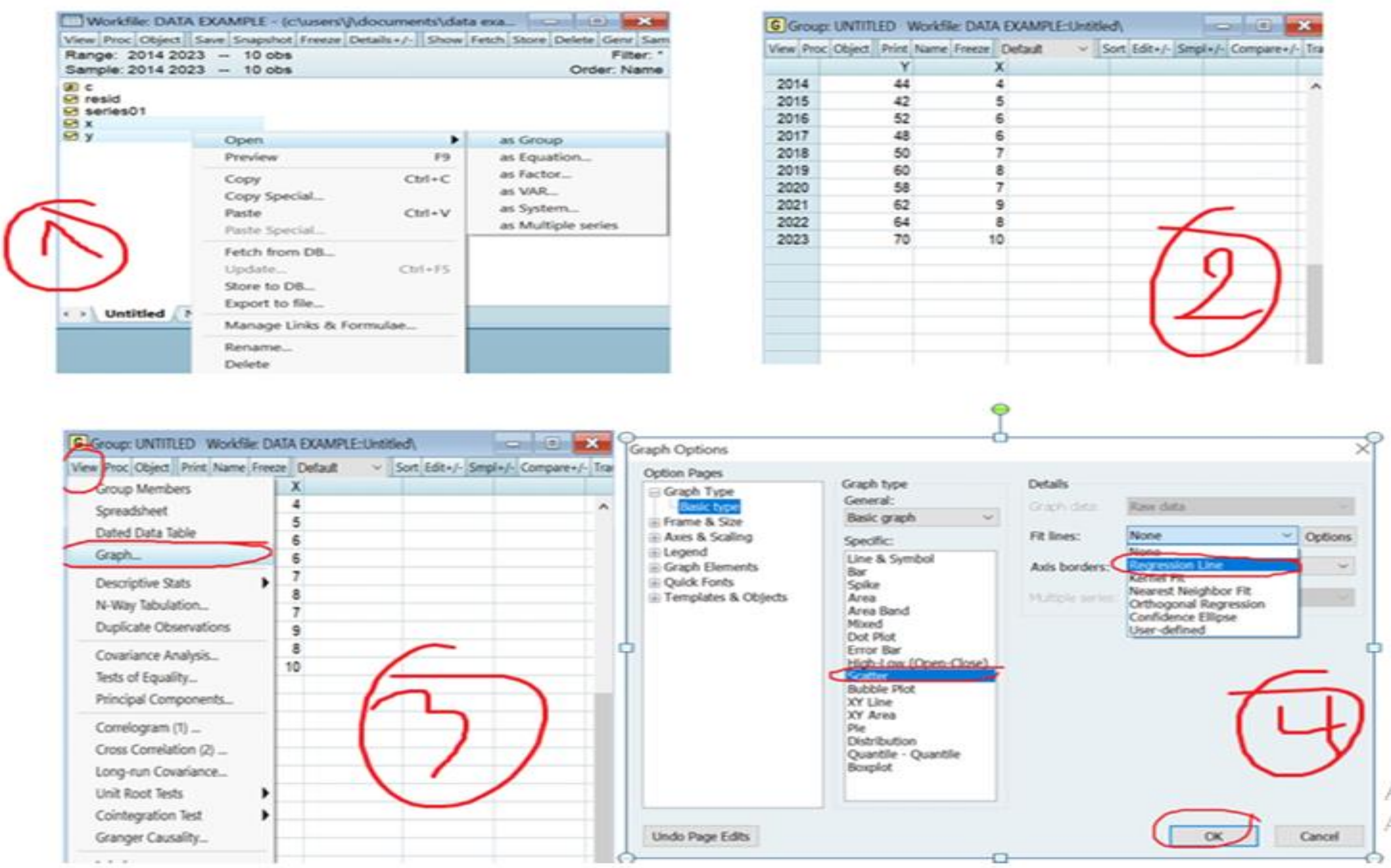
كما تم التطرق إليه في المحاضرة 1، توجد عدة طرق لاستيراد ملف خارجي، إلاّ أننا في هذا التمرين سوف نعتمد على الطريقة الأقصر في استيراد الملف.

File → Import → Import from file → نختار نوع الملف → *ok*

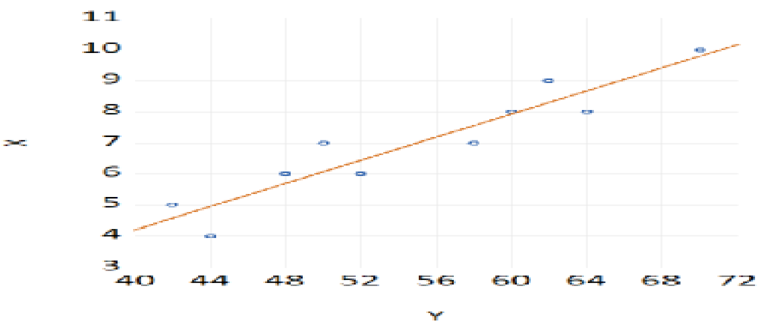
باتباع هذه التعليمات، برنامج EViews يقوم باستيراد الملف مباشرة وإدخال البيانات.

3- العمل على البرنامج:

أ- التمثيل البياني بسحابة النقاط (Scatter) لبيانات الجدول:

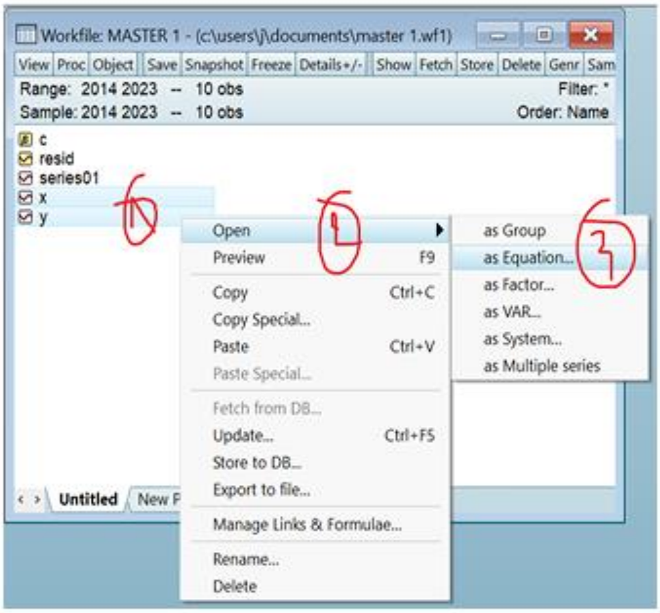


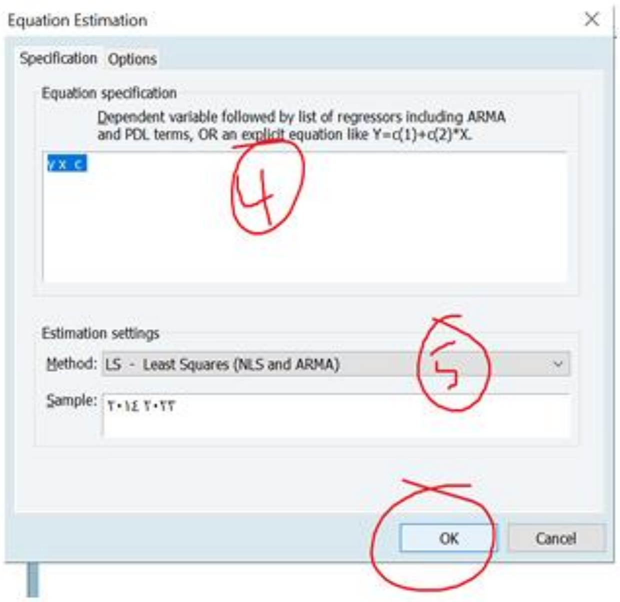
بعد هذه المراحل الأربعة يظهر الشكل البياني المطلوب كالتالي:

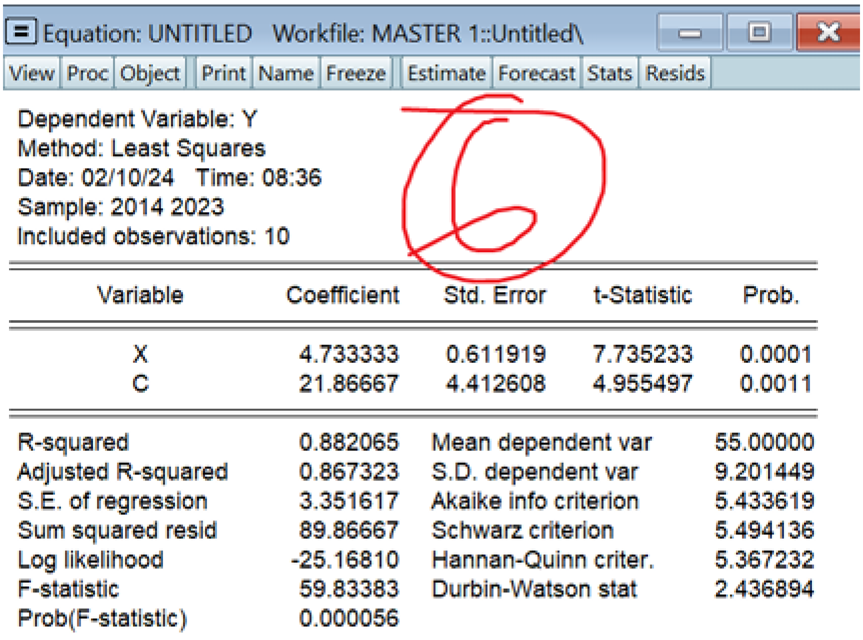


من خلال سحابة النقاط نجد أن العلاقة بين الانفاق على الاعلانات وعوائد مبيعات الشركة علاقة خطية، أي أنه يمكننا التعبير عنها بخط انحدار يمر بين مختلف الثنائيات المشكلة من قيم الانفاق على الاعلانات وقيم عوائد المبيعات، هذا من جهة، ومن جهة أخرى نجد أن هذه العلاقة هي علاقة طرية، بحكم أن زيادة الانفاق على الاعلانات تؤدي إلى زيادة عوائد الأرباح في هذه الشركة. وهو ما سنتأكد منه من خلال إشارة معامل الانحدار الذي سنحسبه لاحقا.

ب- تقدير النموذج وتفسير النتائج:







Dependent Variable: Y
Method: Least Squares
Date: 02/10/24 Time: 08:36
Sample: 2014 2023
Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	4.733333	0.611919	7.735233	0.0001
C	21.86667	4.412608	4.955497	0.0011

R-squared	0.882065	Mean dependent var	55.00000
Adjusted R-squared	0.867323	S.D. dependent var	9.201449
S.E. of regression	3.351617	Akaike info criterion	5.433619
Sum squared resid	89.86667	Schwarz criterion	5.494136
Log likelihood	-25.16810	Hannan-Quinn criter.	5.367232
F-statistic	59.83383	Durbin-Watson stat	2.436894
Prob(F-statistic)	0.000056		

تفسير النتائج:

❖ **من الناحية الاقتصادية:**

- تشير المعلمة $\hat{\alpha} = 21.86$ إلى وجود مقدار ثابت من عوائد المبيعات مقداره 21.86 مليون دج لا يتأثر بتغير الانفاق على الاعلانات.
- تشير المعلمة $\hat{\beta} = +4.73$ إلى وجود علاقة طردية بين الانفاق على الاعلانات وعوائد المبيعات من جهة، وأن كل تغير في الانفاق على الاعلانات بوحدة واحدة يؤدي إلى تغير عوائد المبيعات في نفس الاتجاه بـ 4.73 وحدة.

❖ من الناحية الإحصائية:

- **اختبار ستودنت t- statistic :**

يستعمل هذا الاختبار لدراسة المعنوية الجزئية لمعلمات النموذج عند مستوى معنوية معين، حيث نختبر المعنوية الاحصائية لمعامل الانحدار (β)، والتي تسمح بالحكم على معنوية العلاقة بين المتغير التابع والمتغير المستقل، كما نختبر المعنوية الاحصائية للحد الثابت (α)، والتي تسمح بالحكم على جدوى وجود الحد الثابت في النموذج من عدمها.

بالنسبة لـــ: β

يأخذ اختبار STUDENT الشكل التالي:

$$\begin{cases} H_0 : \beta = 0 \\ H_1 : \beta \neq 0 \end{cases}$$

إحصائيةSTUDENT المحسوبة:

$$St_{cal} = \left| \frac{\hat{\beta}}{\sqrt{\hat{\delta}^2_{\hat{\beta}}}} \right| = \left| \frac{4.73}{0.6119} \right| = 7.73$$

إحصائيةSTUDENT المجدولة:

$$St^{\frac{\alpha}{2}}_{n-2} = St^{\frac{5}{2}}_{10-2} = St^{2.5\%}_{8} = 2.306$$

القرار: نلاحظ أن $St_{cal} \geq St_{tab}$ ، وبالتالي نقبل الفرضية $\beta \neq 0$ ، أي وجود علاقة ذات دلالة احصائية بين الانفاق على الاعلانات وعوائد المبيعات في هذه الشركة. وما يعزز هذا القرار قيمة الاحتمال المرافقة لاحصائية ستودنت Prob = 0.0001 وهي أقل من 5%.

بالنسبة لــ: α

يأخذ اختبار STUDENT الشكل التالي:

$$\begin{cases} H_0 : \alpha = 0 \\ H_1 : \alpha \neq 0 \end{cases}$$

إحصائيةSTUDENT المحسوبة:

$$St_{cal} = \left| \frac{\hat{\alpha}}{\sqrt{\hat{\delta}^2_{\hat{\alpha}}}} \right| = \left| \frac{21.86}{4.4126} \right| = 4.95$$

إحصائيةSTUDENT المجدولة:

$$St^{\frac{\alpha}{2}}_{n-2} = St^{\frac{5}{2}}_{10-2} = St^{2.5\%}_{8} = 2.306$$

القرار: نلاحظ أن $St_{cal} \geq St_{tab}$ ، وبالتالي نقبل الفرضية $\alpha \neq 0$ ، أي وجود الحد الثابت في النموذج له دلالة احصائية. وما يعزز هذا القرار قيمة الاحتمال المرافقة لاحصائية ستودنت Prob = 0.0011 وهي أقل من 5%.

- **اختبار فيشر F- statistic:**

يكون اختبار FISHER عند مستوى معنوية $\alpha = 5\%$ كما يلي:

$$\begin{cases} H_0 : \alpha = \beta = 0 \\ H_1 : \alpha \neq 0 \quad \vee \beta = 0 \end{cases}$$

إحصائيةFISHER المحسوبة، كما يتضح من خلال نتائج التقدير هي:

$$F_{cal} = \frac{R^2 / 2-1}{(1-R^2) / n-2} = \frac{0.8820 / 1}{(1-0.8820) / 10-2} = 59.83$$

إحصائيةFISHER المجدولة:

$$F^{\alpha=5\%}_{(1.n-2)} = F^{\alpha=5\%}_{(1.8)} = 5.318$$

القرار: نلاحظ أن $F_{cal} \geq F_{tab}$ ، وبالتالي نقبل الفرضية $H_1 : \alpha \neq 0 \quad \vee \quad \beta = 0$ ، أي أن النموذج ككل له معنوية إحصائية عند مستوى $\alpha = 5\%$. وما يعزز هذا القرار قيمة الاحتمال المرافقة لاحصائية فيشر Prob(F-Statistic)=0.0000 وهي أقل من 5%.

- **تحليل التباين ومعامل التحديد**

يتكون جدول تحليل التباين من ثلاث أجزاء كما يلي:

$\Sigma(Y_t-\overline{Y})^2$: مجموع مربعات الانحرافات الكلية.(TOTAL SUM OF SQUARES (TSS))

$\Sigma(\hat{Y}_t-\overline{Y})^2$: مجموع مربعات الانحرافات المفسرة.(EXPLAINED SUM OF SQUARES (ESS))

Σe_t^2 : مجموع مربعات البواقي (RESIDUAL SUM OF SQUARES (RSS))

ومنه يمكن إعادة صياغة معادلة تحليل التباين على النحو التالي:

$$TSS = ESS + RSS$$

انطلاقا من معادلة تحليل التباين يمكن استخلاص مؤشر تقاس به القدرة التفسيرية للنموذج، والمتمثل في ما يعرف بمعامل التحديد (R-squared)، والذي نرمز له بالرمز R^2 ، وهو مؤشر يقيس النسبة المفسرة من التغير الكلي بدلالة خط الانحدار، أي نسبة مجموع مربعات الانحرافات المفسرة إلى مجموع مربعات الانحرافات الكلية، تتراوح قيمته بين الصفر والواحد، أي: $0 \le R^2 \le 1$ ، فكلما اقترب R^2 من الواحد، تكون للنموذج قدرة تفسيرية عالية، وكلما اقترب من الصفر، دل ذلك على ضعف القدرة التفسيرية للنموذج. بصيغة رياضية يكتب معامل التحديد كما يلي:

$$R^2 = \frac{ESS}{TSS} = \frac{\Sigma(\hat{Y}_t-\overline{Y})^2}{\Sigma(Y_t-\overline{Y})^2}$$

أو بطريقة أخرى:

$$R^2 = 1-\frac{RSS}{TSS} = 1-\frac{\Sigma e_t^2}{\Sigma(Y_t-\overline{Y})^2}$$

أما جدول تحليل التباين (ANOVA) Analysis Of Variance فيأخذ الشكل التالي:

مصدر التغير	مجموع المربعات	درجة الحرية	متوسط المربعات
المتغير المستقل X_t	$ESS=\Sigma(\hat{Y}_t-\overline{Y})^2$	1	$\Sigma(\hat{Y}_t-\overline{Y})^2/1$
البواقي e_t	$RSS=\Sigma e_t^2$	$n-2$	$\Sigma e_t^2/n-2$
المجموع	$TSS=\Sigma(Y_t-\overline{Y})^2$	$n-1$	

من خلال نتائج برنامج EViews نجد:

$$R^2 = 1-\frac{\Sigma e_t^2}{\Sigma(Y_t-\overline{Y})^2} = 1-\frac{89.86}{762} = 1-0.1179 = 0.8820$$

بعبارة أخرى، معامل التحديد يحسب مباشرة من برنامج إفيوز بالصيغة التالية:

$$\mathbf{R-squared} = 1-\frac{Sum\ squqred\ resid}{(S.D.dependent\ var)^2.(n-1)} = 1-\frac{89.86667}{(9.201449)^2.(10-1)} = 1-\frac{89.86667}{762} = \mathbf{0.8820}$$

ت- حساب القيم المقدرة $\hat{Y}_t$ واستنتاج بواقي التقدير e_t:

$$\hat{Y}_t = 21.86 + 4.73\cdot X_t$$

$$\hat{Y}_{2014} = 21.86+4.73\cdot X_{2014} = 21.86+4.73(4) = 40.80 \quad \Rightarrow e_{2014} = Y_{2014}-\hat{Y}_{2014} = 44-40.80 = +03.20$$

$$\hat{Y}_{2015} = 21.86+4.73\cdot X_{2015} = 21.86+4.73(5) = 45.53 \quad \Rightarrow e_{2015} = Y_{2015}-\hat{Y}_{2015} = 42-45.53 = -03.53$$

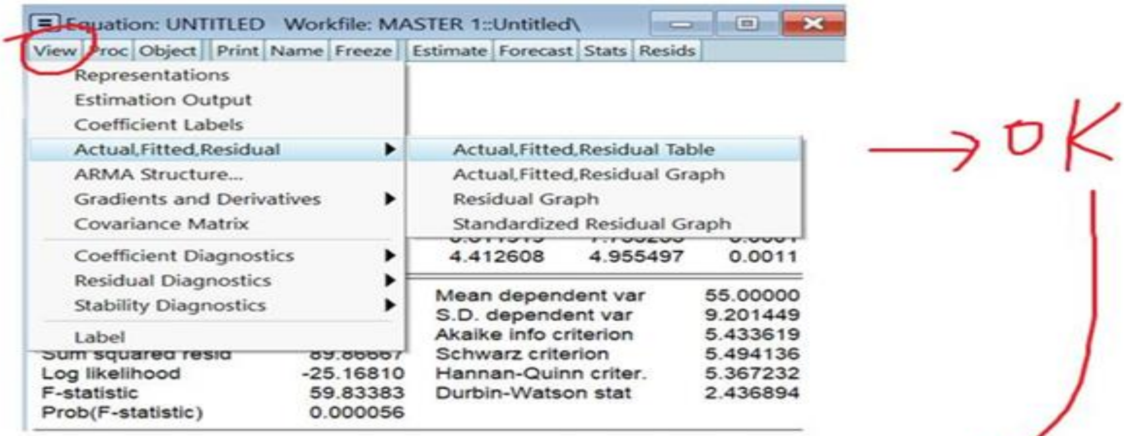
$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$

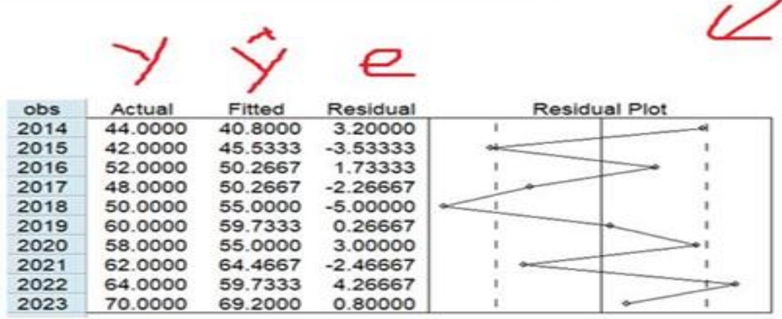
$$\hat{Y}_{2023} = 21.86+4.73\cdot X_{2023} = 21.86+4.73(10) = 69.19 \quad \Rightarrow e_{2023} = Y_{2023}-\hat{Y}_{2023} = 70-69.19 = +0.81$$

االليبليل

باستخدام برنامج EViews يتم استخراج القيم المتوقعة للمتغير التابع وبواقي التقدير بتطبيق التعليمات التالية:

$$Table\ Estimation\ \rightarrow View \rightarrow Actual, Fitted, Residual \rightarrow Actual, Fitted, Residual\ Table \rightarrow ok.$$





	y	$\hat{y}$	e	
obs	Actual	Fitted	Residual	Residual Plot
2014	44.0000	40.8000	3.20000	
2015	42.0000	45.5333	-3.53333	
2016	52.0000	50.2667	1.73333	
2017	48.0000	50.2667	-2.26667	
2018	50.0000	55.0000	-5.00000	
2019	60.0000	59.7333	0.26667	
2020	58.0000	55.0000	3.00000	
2021	62.0000	64.4667	-2.46667	
2022	64.0000	59.7333	4.26667	
2023	70.0000	69.2000	0.80000	