تمهيد:

يعتبر الانحدار الخطي أحد الأساليب الإحصائية التي تستخدم في قياس العلاقات الاقتصادية بين متغير يسمى المتغير التابع وآخر أو مجموعة من المتغيرات تسمى المتغيرات المستقلة أو التفسيرية، والذي يعتبر تحديدها إحدى مسؤوليات الباحث وذلك باستناده على مصادر عدة لعل أبرزها النظريات الاقتصادية والتي تعتبر خير مرشد سواء كانت كلية والمرتبطة بالظواهر الكلية مثال ذلك النظرية النقدية الكينزية التي تعتبر أن الطلب على النقود كمتغير تابع وأسعار الفائدة والدخل النقدي الجاري كمتغيرات مفسرة.

أو كانت جزئية تفسر سلوك الوحدات الاقتصادية نأخذ عل سبيل المثال الكمية المطلوبة من سلعة معينة كمتغير تابع تتأثر بمجموعة من المتغيرات المفسرة مثل الدخل، السعر، أسعار السلع البديلة، أسعار السلع المكملة.

وتنقسم نماذج الانحدار الخطي إلى نوعين الانحدار الخطي البسيط والانحدار الخطي المتعدد، وتتحدد نماذج الانحدار الخطي بعدد المتغيرات المستقلة حيث يدرس الانحدار الخطي البسيط العلاقة بين متغير تابع وآخر مستقل في حين الانحدار المتعدد يدرس العلاقة بين متغير تابع وعدة متغيرات مستقلة.

1-مفهوم الانحدار الخطي البسيط:

يعتبر الانحدار البسيط من أبسط نماذج الانحدار المختلفة للتقدير والتحليل الإحصائي والاقتصادي، إذ يقتصر على وصف علاقة خطية عشوائية تربط بين متغيرين أحدهما تابع والآخر مستقل، وتشكل دراسة الانحدار البسيط القاعدة الأساسية ونقطة الانطلاق نحو دراسة نماذج أخرى أكثر عمومية وواقعية، بحيث توجد العديد من العلاقات الاقتصادية التي يمكن قياسها باستخدام هذا الأسلوب، على سبيل المثال النظرية النقدية الكلاسيكية التي تمت دراستها على مستوى الفصل الأول تعتبر أن كمية النقود أو العرض النقدي متغير مفسر والمستوى العام للأسعار المتغير التابع $p = f(M)$.

وتأخذ معادلة الانحدار البسيط الصيغة الخطية التالية:

$Y_i = B_0 + B_1 X_i$ ..... (01) حيث يمثل (Y_i) المتغير التابع و (X_i) المتغير المستقل أو التفسيري الذي يفسر تغيرات المتغير التابع، في حين تمثل B_0 و B_1 معلمات النموذج.

إلا أنه ما يلاحظ على المعادلة رقم(01) أنها لا تعبر عن حقيقة العلاقة بين المتغيرين التابع والمستقل وذلك لأنه إذا ما تم الرجوع إلى الواقع لا يوجد أثر لهذه العلاقات الدقيقة، فالبيانات التي يتم استسقاءها من الواقع خلال فترات معينة تعطينا شكلا انتشاريا مبعثرا لا تتجمع نقاطه على خط مستقيم مما يوحي بوجود انحراف للعلاقة الحقيقية عن العلاقة المعبر عنها بالمعادلة، ولتقريبها من الواقع والحقيقة ينبغي إضافة متغير آخر يسمى المتغير العشوائي أو حد الخطأ ويرمز له بالرمز U_i.

فتصبح المعادلة السابقة التي تعبر عن حقيقة العلاقة والتي تسمى معادلة الانحدار الخطي البسيط كالآتي:

$$Y = B_0 + B_1 X_i + U_i$$

عندها تصبح الصيغة القياسية جاهزة للتقدير وقريبة من الواقع.

ويستخدم مصطلح المتغير العشوائي للتعبير عن حجم الخطأ أو عدم إصابة الهدف وهناك عدة أسباب تبرر إدخال هذا المتغير ونجملها في:

1-صعوبة إدخال كافة المتغيرات المؤثرة في الظاهرة: ويقصد بهذا العنصر صعوبة الحصول على كافة البيانات المتعلقة بالمتغيرات التي تؤثر في الظاهرة محل الدراسة لهذا يكون من الضروري إدخال عنصر ينوب عنها وهو العنصر العشوائي.

2-صعوبة إدخال المتغيرات غير المتوقعة: قد يتم إضافة المتغير العشوائي إلى المعادلة لكي يعبر عن المتغيرات غير المتوقعة والتي تظهر بالإضافة إلى العناصر الأخرى المؤثرة في المتغير التابع، كالحروب والبراكين، الكوارث الطبيعية الأخرى، أو بعض المتغيرات التي لها تأثير قليل على المتغير التابع وذات مؤشرات ضعيفة لا يمكن تقديرها.

3-الخطأ في القياس: في كثير من الأحيان يحدث خطأ في قياس المتغيرات الاقتصادية فيتم إدراج المتغير العشوائي في المعادلة.

4-التصرفات العشوائية غير المتوقعة للأفراد: يختلف سلوك الأفراد عشوائيا حتى ولو كانت الظروف المحيطة بهم متشابهة، فيتم تضمين المعادلة المتغير العشوائي ليمثل العشوائية المتأصلة في سلوك الأفراد، فليس بالضرورة أن تتفق أسر متشابهة من حيث الدخل وعدد أفرادها وأعمارهم على نفس المقدار من الاستهلاك.

5-عدم الصياغة الدقيقة للنموذج: ويعني ذلك وقوع خطأ في تخصيص الشكل الدالي للعلاقة المدروسة فيتم إدخال المتغير العشوائي.

2-تحديد الشكل الرياضي للنموذج الخطي البسيط وفرضياته:

2-1-الشكل الرياضي للنموذج الخطي البسيط: ويقصد بذلك وصف العلاقة الخطية بين المتغيرين (X_i) و(Y_i) لعينةnمن المشاهدات بواسطة نموذج قياسي في شكل رياضي وليكن كما يلي:

$$Y_i = B_0 + B_1 X_i + U_i \quad ، \quad (i = 1.....n)$$

حيث:

Y_i: يسمى المتغير التابع أو المتغير الداخلي، يمكن مشاهدته.

X_i: يسمى المتغير المفسر أو المتغير الخارجي، يمكن مشاهدته.

U_i: يسمى الحد العشوائي أو حد الخطأ لا يمكن مشاهدته فهو يشير إلى أن المتغير المفسر X_i لا يمكن لوحده تفسير كل التغيرات التي قد تطرأ على المغير التابعY_i، بل هناك عوامل أخرى لها علاقة بما يحدث لـ Y_iفينوب عنها.

أما القيمB_0 وB_1 فتمثل معلمات النموذج وهي عبارة عن قيم ثابتة وغير معلومة في المعادلة.

ونتيجة لوجود المتغير العشوائي في المعادلة السابقة فإنه يمكن تقسيم العلاقة أعلاه إلى جزئين:

- الجزء الأول يتضمن الخط($B_0 + B_1 X_i$)والتي تعرف بتغيرات المتغير التوضيحي.

- الجزء الثاني فيتمثل في التأثير العشوائي(U_i) وتعرف بالتغيرات غير الموضحة.

كما يمكن التعبير عن العلاقة الخطية للمعادلة السابقة في الشكل المصفوفي الآتي:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}, \quad U = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix}$$

قبل التطرق إلى فرضيات نموذج الانحدار الخطي البسيط يجدر التذكير بأنه تتعدد الصيغ الرياضية التي تعبر عن العلاقة بين المتغير التابع والمتغير المستقل من خطية وغير خطية التي يمكن أن يصادفها باحث القياس الاقتصادي، ويمكن تلخيص أهم الصيغ في الجدول التالي:

جدول رقم (01): يبين الصيغ الرياضية

نوع الدالة	الصيغة غير الخطية	الصيغة الخطية
الصيغة الخطية	....	$Y_{ij} = B_0 + B_1 X_j + U_{ij}$
الصيغة العكسية	....	$Y_{ij} = B_0 - B_1/X_j + U_{ij}$
الصيغة التربيعية	....	$Y = B_0 + B_1 X + B_2 X^2$
الصيغة نصف اللوغاريتمية	$e^Y = e^{B_0} X^{B_1}$	$Y_{ij} = B_0 + B_1 \ln X_j + U_{ij}$
الصيغة اللوغاريتمية المزدوجة	$Y = B_0 + X^{B_1}$	$\ln Y_{ij} = B_0 + B_1 \ln X_j + U_{ij}$
الصيغة الأسية	$Y = e^{B_1 + B_1 X}$	$\ln Y = B_0 + B_1 X$

المصدر: مجدي الشوربجي، الاقتصاد القياسي النظرية والتطبيق، الطبعة الأولى، الدار المصرية اللبنانية، القاهرة، 1994، ص. 49 بتصرف.

ولأن الصيغة الخطية هي التي يتم استخدامها في التقدير، فيكون من الضروري على باحث القياس الاقتصادي تحويل الصيغ الرياضية غير الخطية بإدخال اللوغاريتم الطبيعي إلى صيغة خطية وذلك حتى يمكن إجراء عملية التقدير بدقة.

2-2 -فرضيات النموذج: من أجل تقدير معلمات النموذج لابد من الاستناد إلى قاعدة سليمة تتمثل في مجموعة من الفروض الأساسية والتي يمكن تلخيصها في القائمة التالية:

الفرضية الأول: تتعلق بقيم المتغير المستقل، حيث يفترض أن البيانات التي جمعت بالنسبة لهذا المتغير قادرة على إظهار تأثيراتها في تغير قيم المتغير التابع، وتكون قيمة واحدة على الأقل من قيم المتغير المستقل مختلفة عن بقية القيم ويمكن صياغة هذا الفرض بالصورة الرياضية التالية:

$$\sum (X_i - \bar{X})^2 \neq 0$$

حيث أن $(\bar{X})$ تعبر عن متوسط قيم(X_i).

الفرضية الثانية: الأمل الرياضي للأخطاء معدوم أي أن:

$$(E(U_i) = 0) \quad \forall(i = 1.....n)$$

وتعني هذه الفرضية أن الأخطاء لا تدخل في تفسير (Y_i)، إذ تعبر عن حدود عشوائية تأخذ قيما سالبة، موجبة أو معدومة لا يمكن قياسها أو تحديدها بدقة، تخضع لقوانين الاحتمال بحيث يكون وسطها أو توقعها الرياضي مساو للصفر.

الفرضية الثالثة: وتتمثل في أن يكون تباين حد الخطأ متساوي عند جميع قيم (X_i)ويبلغ (σ^2)وهذه القيمة مجهولة ويعبر عنها رياضيا كما يلي:

$$VAR\ (U_i) = E\ (U_i{}^2) = \sigma^2$$

أي أن قيمة التباين ثابتة عند قيم(X_i) المختلفة، مما يعني أن البيانات التي جمعت لتقدير العلاقة يمكن الاعتماد عليها بنفس الدرجة، فكل مشاهدة تؤثر بنفس القوة في العلاقة التي يطلب تقديرها.

ومن الفرض الثاني والفرض الثالث نستنتج أن المتغير العشوائي يتوزع توزيعا طبيعيا أي:

$$U_i \rightarrow N(0, \sigma^2)$$

الفرضية الرابعة: تتعلق بالارتباط الذاتي بالمتغير العشوائي، ويعني أن تكون قيمة (U_i)مستقلة عن (U_S)و$(i \neq s)$ أي أن القيم المختلفة للمتغير العشوائي (U_i)تكون مستقلة عن بعضها البعض، بمعنى التباين المشترك(U_i) مع (U_S)يساوي الصفر ويمكن كتابة هذه الفرضية رياضيا كما يلي:

$$Cov\ (U_iU_S) = E\ (U_iU_s) = 0 \quad ، \quad \forall\ i \neq S\ ،\ (\ i\ ,\ S)\ = 1.....n$$

ويعود السبب في ذلك إلى كوننا نريد أن نصف نموذجا تكون فيه قوة واحدة فقط تؤثر في قدرته على التنبؤ، فإذا افترضنا أن (U_i)ترتبط بقيمة(U_S) فتكون قيم(Y_i)لا تعتمد فقط على قيمة (X_i)وإنما تعتمد بالإضافة إلى ذلك على(U_S) أي القيم السابقة واللاحقة لها، وبذلك تكون أكثر من قوة تؤثر في قدرة النموذج على التنبؤ.

الفرضية الخامسة: تتمثل هذه الفرضية في أن تكون قيم(U_i) غير مرتبطة بأي من المتغيرات المستقلة في النموذج، ويعبر عنه رياضيا كما يلي:

$$E(U_iX_i) = 0$$

أما فيما يخص الفروض التي تتعلق بالمتغير التابع الممثل للظاهرة:

- أن يكون توزيع هذا المتغير التابع توزيع طبيعي وتوقعه أو متوسطه يعطى كالآتي:

$$\bar{Y} = E(Y_i) = B_0 + B_1X_i$$

- وتباينه يعطى بالشكل التالي:

$$Var(Y_i) = E[(Y_i - E(Y_i)]^2 = \sigma^2$$

بمعنى يكون المتغير التابع خاضع للقانون التالي:

$$Y_i \to N(B_0 + B_1 X_i \ , \ \sigma^2)$$

وبتوفر هذه الفرضيات يمكن تقدير معلمات النموذج.

3-تقدير معلمات النموذج:

انطلاقا من الصيغة الخطية للنموذج البسيط:

$$Y_i = B_0 + B_1 X_i + U_i$$

ومن القانون الاحتمالي الذي يخضع له المتغير العشوائي $U_i \to N(0, \sigma^2)$

فإن المعلمات المجهولة والتي ينبغي تقديرها من النموذج $B_0 \ , B_1 \ , \sigma^2$.

فيعرف المقدر بأنه الطريقة أو الصيغة التي تسمح بتقدير المعلمات غير المعلومة، أما التقدير فهو القيمة العددية الناتجة عن تطبيق تلك الصيغة أو الطريقة على بيانات العينة، فتكون المعادلة التقديرية على الشكل التالي:

$$\hat{Y} = \hat{B}_0 + \hat{B}_1 X$$

وتتعدد طرق تقدير معلمات النموذج المجهولة، إلا أن أحسنها طريقة المربعات الصغرى العادية التي اعتمد عليها لأول مرة العالم الألماني كارل فريدريك جاوس(Gauss) عام 1821، وذلك لما تتميز به من خصائص إحصائية وسهولة الحساب، كما أنها طريقة تقدير طبيعية تتوافق مع هيكل النموذج، فضلا على أن مقدراتها أفضل مقدر خطي أمثل حتى وإن كان النموذج الحقيقي ليس انحدار خطي، كما تتميز بالدقة التي تنعدم لدى غيرها من المقدرات لذلك تعتبر الاستعمال الفعال للبيانات.

1-3 -طريقة المربعات الصغرى: تعد طريقة المربعات الصغرى المحرك الأساسي في التحليل الإحصائي وأحسن أسلوب يتم الاعتماد عليه في اختيار أفضل خط مستقيم يوفق بين مشاهدات العينة من بين الخطوط اللانهائية العدد التي تصف المعادلة الخطية تقوم على مبدأ أساسي يتمثل في تصغير مجموع مربعات الأخطاء إلى أقل قدر ممكن، فعند سحب عينة مكونة من n مشاهدة من الثنائيات (X_i, Y_i) يكون الغرض هو انتقاء أفضل خط مستقيم يعبر عن العلاقة $Y_i = B_0 + B_1 X_i + U_i$ ويكون تقديرها بالصيغة التالية:

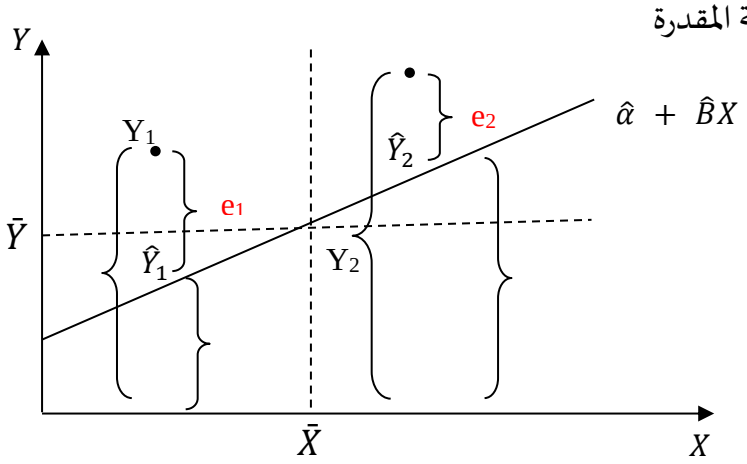
$$\hat{Y}_i = \hat{B}_0 + \hat{B}_1 X_i$$

حيث:

$\hat{Y}_i$: يمثل القيمة التقديرية للقيمة الحقيقية لـY_i.

وعند تمثيل بيانات العينة نتحصل على سحابة من النقاط منتشرة في شكل انتشاري كما يوضحه الرسم البياني التالي:

شكل رقم (02): يبين العلاقة المقدرة



المصدر: عبد المحمود محمد عبد الرحمن، مرجع سبق ذكره، ص. 12

يتبين من خلال الشكل أعلاه أن النقاط تتوزع أعلى وأسفل الخط التقديري، تمثل سلسلة الملاحظات الحقيقية Y_i تنحرف عن سلسلة الملاحظات التقديرية $\hat{Y}_i$ (بقيم سالبة أو موجبة أو معدومة)، وتسمى هذه الانحرافات المبينة في الشكل بالبواقي (بواقي المربعات الصغرى) أو مقدرات الأخطاء والتي نرمز لها بالرمز e_i حيث:

$$e_i = Y_i - \hat{Y}_i$$

وتقوم طريقة المربعات الصغرى على المبدأ الأول وهو تصغير مجموع البواقي إلى أدنى حد ممكن أو جعله معدوم أي $\sum e_i = 0$ لتعطي أفضل خط مستقيم يوفق مشاهدات العينة (X_i, Y_i) إلا أن هذا الشرط غير كاف وذلك لأنه يمكن للبواقي السالبة أن تلغي البواقي الموجبة وتكون النتيجة صفر، دون أن يتحقق شرط اختيار أفضل مقدرتين B_0 ، B_1 لذا تشترط طريقة المربعات الصغرى قيد آخر يتمثل في تصغير مجموع مربعات البواقي إلى أدنى قيمة ممكنة لتصبح كل البواقي مربعة وبالتالي موجبة، أي $\sum {e_i}^2$ ويتم التعبير عنها رياضيا كما يلي:

$$Q = \sum_{i=1}^{n} {e_i}^2 = \sum (Y_i - \hat{B}_0 - \hat{B}_1 X_i)^2$$

وبالتفاضل الجزئي للدالة Q بالنسبة لـ $\hat{B}_0$ ، $\hat{B}_1$ ومساواتها إلى الصفر نجد:

$$\begin{cases} \dfrac{\delta(Q)}{\delta(\hat{B}_0)} = -2\sum (Y_i - \hat{B}_0 - \hat{B}_1 X_i) = 0 \\ \dfrac{\delta(Q)}{\delta(\hat{B}_1)} = -2\sum X_i (Y_i - \hat{B}_0 - \hat{B}_1 X_i) = 0 \end{cases}$$

وبتبسيط المعادلتين نتوصل إلى ما يسمى بالمعادلات الطبيعية الخاصة بـ $\hat{B}_1$ ، $\hat{B}_0$ للخط المستقيم وبحل جملة المعادلتين نجد قيم كل من $\hat{B}_1$ ، $\hat{B}_0$ أي أن:

$$\begin{cases} \hat{B}_1 = \dfrac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} \\ \hat{B}_0 = \dfrac{\sum {X_i}^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n\sum {X_i}^2 - (\sum {X_i}^2)} \\ \quad = \bar{Y} - \hat{B}_1 \bar{X} \end{cases}$$

ويعرف هذا التقدير بالتقدير حول نقطة الأصـــل، أما التقدير حول نقطة المتوســـط فيمكن كتابته على الشكل التالي:

$$x_i = X_i - \bar{X} \quad ، \quad y_i = Y_i - \bar{Y}$$

وعليه يمكن إعادة كتابة الصيغة الخاصة $\hat{B}_1$ بالشكل التالي $\hat{B}_1 = \frac{\sum x_i y_i}{\sum {x_i}^2}$، أما الحد الثابت فيقدر بنفس الصيغة السابقة[1].

وهكذا إذا ما تم الحصـــول على تقديرات المربعات الصـــغرى فإنه يمكن تحديد خط الانحدار المقدر والخاص بالعينة حيث يعطى بالمعادلة التالية:

$$Y_i = \hat{B}_0 + \hat{B}_1 X_i + U_i \quad \text{أو} \quad \hat{Y} = \hat{B}_0 + \hat{B}_1 X_i$$

مثال رقم(01): البيانات التالية تمثل الدخل الشهري X والانفاق الشهري Y للعائلة تتكون من 9 عائلات كما يلي:

X	60	90	56	18	42	38	90	80	48
Y	40	55	38	12	35	30	72	60	45

المطلوب:

-ارسم الشكل الانتشاري للبيانات.

-قدر معادلة الانحدار الخطي البسيط.

[1] أموري هادي كاظم الحسناوي، مرجع سبق ذكره، ص. 17-18