تحليل الانحدار الخطي المتعدد

في الواقع الاقتصادي، لا يمكن الاستعانة بالنموذج ذي متغيرين لتحليل الظاهرة الاقتصادية حيث أن هذه الأخيرة لا تفسَّر فقط بمحدد واحد و إنما ينبغي إدماج جميع المحددات أو العوامل المؤثرة في الظاهرة لكي تكون الدراسة أكثر شمولية. في هذا الفصل، نقوم بدراسة الانحدار العام و ذلك باقتراح طريقة لتقدير معالم النموذج و دراسة الخصائص الإحصائية للمقدرات ثم اختبار الفرضيات.

1. الصياغة الرياضية للنموذج الخطي العام :

يستند النموذج الخطي العام على افتراض وجود علاقة خطية ما بين متغير معتمد Y_i وعدد من المتغيرات المستقلة:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + \varepsilon_i , \quad i = 1, \ldots\ldots, n$$

المتغيرات $X_{i1}, \ldots, X_{ij}, \ldots, X_{ik}$ تسمى المتغيرات المُفَسِّرة أو المستقلة للمتغير المفسَّر أو التابع Y_i وما يجب ملاحظته أن Y_i مشروح من طرف k متغير مُفَسِّر و لا يمكن لهذه الأخيرة أن تفسر Y بشكل تام، لأنه لا يمكننا في غالب الأحيان حصر جميع الظواهر المؤثرة على Y (بعض الظواهر غير قابلة للتكميم)، لذلك يُدرج حد الخطأ ε_i الذي يتضمن كل المعلومات التي لا تقدمها المتغيرات المفسرة و نفترض عادة بأن المتغيرات المستقلة كلما أخذت بعين الاعتبار كلما كانت المعلومات التي يقدمها الخطأ العشوائي مهملة. نشير فقط إلى أن $\beta_0, \beta_1, \ldots., \beta_k$ هي معالم النموذج، لدينا هنا $(k+1)$ معلم في النموذج.

الـ n مشاهدة تعطينا n معادلة :

$$i=1: Y_1=\beta_0+\beta_1 X_{11}+\beta_2 X_{12}+.....+\beta_k X_{1k}+\varepsilon_1$$
$$i=2: Y_2=\beta_0+\beta_1 X_{21}+\beta_2 X_{22}+.....+\beta_k X_{2k}+\varepsilon_2$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$i=n: Y_n=\beta_0+\beta_1 X_{n1}+\beta_2 X_{n2}+.....+\beta_k X_{nk}+\varepsilon_n$$

يمكن كتابة هذا النظام على الشكل المصفوفي التالي: $Y = X\beta+\varepsilon$

$$Y=\begin{pmatrix} Y_1\\ Y_2\\ \vdots\\ \vdots\\ Y_n\end{pmatrix},\quad X=\begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k}\\ 1 & X_{21} & X_{22} & \cdots & X_{2k}\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk}\end{pmatrix},\quad \beta=\begin{pmatrix} \beta_0\\ \beta_1\\ \beta_2\\ \vdots\\ \beta_k\end{pmatrix},\quad \varepsilon=\begin{pmatrix} \varepsilon_1\\ \varepsilon_2\\ \vdots\\ \vdots\\ \varepsilon_n\end{pmatrix}$$

$Y(n\times1)$: المتغير التابع أو المفسَّر،

$X(n\times(k+1))$: مصفوفة المتغيرات المُفَسِّرة أو المستقلة،

$\beta((k+1)\times1)$: شعاع المعالم،

$\varepsilon(n\times1)$: شعاع الأخطاء.

2. الفرضيات الأساسية للنموذج:

إن بناء نموذج الانحدار الخطي يجب أن يكون مستوفيا لعدد من الفرضيات التي يمكن إجمالها كما يلي:

- الفرضية الأولى: المتغيرات المُفَسِّرة المهملة في النموذج لها أثر متوسط معدوم $E(\varepsilon)=0$.
- الفرضية الثانية :

$$\begin{cases} \operatorname{var}(\varepsilon_i) = \sigma^2, & \forall i = 1.........n \\ \operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0, & \forall i \neq j \end{cases}$$

حيث أن $\operatorname{var}(\varepsilon_i) = \sigma^2 \quad \forall i = 1.........n$ هي فرضية تجانس التباين "Homoscedasticity" لمختلف الحدود العشوائية، وهذا كفيل بإبعاد الحالة التي تكون فيها الأخطاء تتبع تغيرات قيم المتغيرات المفسرة و $Cov(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i \neq j$، أي أن الأخطاء ليست مرتبطة ببعضها، وأن نتيجة تجربة لا تؤثر على بقية النتائج. يمكن كتابة هاتين الفرضيتين على الشكل المصفوفي :

$$\Omega_\varepsilon = E(\varepsilon\varepsilon') = \begin{pmatrix} \sigma_\varepsilon^2 & 0 & \cdots & 0 \\ 0 & \sigma_\varepsilon^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_\varepsilon^2 \end{pmatrix} = \sigma_\varepsilon^2 I_n$$

تسمى المصفوفة Ω_ε مصفوفة التباين– التباين المشترك للأخطاء.

- ❖ الفرضية الثالثة: المصفوفة X غير عشوائية وثابتة: تعني بأن قيم المتغيرات المستقلة يمكن مراقبتها، وبالإضافة إلى ذلك نفترض X ثابتة لضمان بأن قيم المتغيرات المستقلة لا تتغير من حين لآخر، أي ;

$$\operatorname{cov}(X, \varepsilon) = E(X'\varepsilon) = 0$$

- ❖ الفرضية الرابعة: عدد المشاهدات n هو أكبر من عدد المتغيرات المفسرة k ، وهي الحالة التي تلغي الارتباط الخطي بين المتغيرات المستقلة.

3. تقدير شعاع المعالم β وتباين الأخطاء σ^2:

علينا تقدير βبشكل يجعل $\hat{Y}$ أقرب ما يمكن للمتغير التابع Y، ولهذا الغرض توجد عدة طرق، على غرار طريقة المربعات الصغرى.

1.3. طريقة المربعات الصغرى :

1.1.3. تقدير شعاع المعالم β:

تهدف هذه الطريقة إلى إيجاد تقدير للشعاع β الذي يعمل على تدنئة مجموع مربعات الأخطاء $\hat{\varepsilon}_i$ بين القيمة المقدرة $\hat{Y}$ والقيمة الحقيقية Y. $Min\sum_{i=1}^{n}\hat{\varepsilon}_i^2 = Min\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$

بعد عمليات التبسيط ينتج: $\boxed{\hat{\beta} = (X'X)^{-1}X'Y}$

2.1.3. تقدير تباين الأخطاء σ^2 و مصفوفة التباين–التباين المشترك للمقدرات $\Omega_{\hat{\beta}}$:

$$\boxed{\hat{\sigma}_\varepsilon^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k-1} = \frac{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{n-k-1}}$$

بما أن σ_ε^2 غير معروف، فانه يمكن استبداله بمقدر تباين الأخطاء $\hat{\sigma}_\varepsilon^2$ وعليه تكون مصفوفة التباين – التباين المشترك للمقدرات من الشكل: :

$$\hat{\Omega}_{\hat{\beta}} = \hat{\sigma}_\varepsilon^2 (X'X)^{-1}$$

<u>مثال 1:</u>

ليكن لدينا البيانات الإحصائية التالية: المتغير التابع Y، المتغيرين المستقليين X_1 و X_2

Y_i	X_{i1}	X_{i2}
4	2	3
6	4	7
7	5	10
9	7	8
10	9	8
12	10	9
14	12	11
16	14	13
18	15	14
20	17	15

المطلوب:

1. اكتب النموذج على الشكل المصفوفي

2. أوجد شعاع معالم النموذج

3. مصفوفة التباين – التباين المشترك للمقدرات

الحل:

1. الكتابة على الشكل المصفوفي $Y = X\beta + \varepsilon$

$$\begin{pmatrix} 4 \\ 6 \\ 7 \\ \vdots \\ 20 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 7 \\ 1 & 5 & 10 \\ \vdots & \vdots & \vdots \\ 1 & 17 & 15 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_{10} \end{pmatrix}$$

2. نعلم أن $\hat{\beta} = (X'X)^{-1} X'Y$. نقوم إذن بحساب $X'X$ و $(X'X)^{-1}$ ثم $X'Y$

$$X'X = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 2 & 4 & 5 & \cdots & 17 \\ 3 & 7 & 10 & \cdots & 15 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 7 \\ 1 & 5 & 10 \\ \vdots & \vdots & \vdots \\ 1 & 17 & 15 \end{pmatrix} = \begin{pmatrix} 10 & 95 & 98 \\ 95 & 1129 & 1081 \\ 98 & 1081 & 1078 \end{pmatrix}$$

و بعد حساب المعكوس نحصل على:

$$(X'X)^{-1} = \begin{pmatrix} 1.1725 & 0.0852 & -0.1921 \\ 0.0852 & 0.0284 & -0.0362 \\ -0.1921 & -0.0362 & 0.0547 \end{pmatrix}$$

و أيضا:

$$X'Y = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 2 & 4 & 5 & \cdots & 17 \\ 3 & 7 & 10 & \cdots & 15 \end{pmatrix} \begin{pmatrix} 4 \\ 6 \\ 7 \\ \vdots \\ 20 \end{pmatrix} = \begin{pmatrix} 116 \\ 1342 \\ 1298 \end{pmatrix}$$

وعليه يكون :

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X'X)^{-1} X'Y = \begin{pmatrix} 1.1725 & 0.0852 & -0.1921 \\ 0.0852 & 0.0284 & -0.0362 \\ -0.1921 & -0.0362 & 0.0547 \end{pmatrix} \begin{pmatrix} 116 \\ 1342 \\ 1298 \end{pmatrix} = \begin{pmatrix} 1.0996 \\ 0.9776 \\ 0.1237 \end{pmatrix}$$

لإيجاد مصفوفة التباين–التباين المشترك للمقدرات، ينبغي أولا حساب تباين البواقي حيث يتم حساب قيم $\hat{Y}_i$ انطلاقا من الانحدار الخطي $\hat{Y}_i = 1.0996 + 0.9776X_{i1} + 0.1237X_{i2}$ و بواقي التقدير $\hat{\varepsilon}_i$ من المعادلة $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$.

حساب البواقي

Y_i	$\hat{Y}_i$	$\hat{\varepsilon}_i$	$\hat{\varepsilon}_i^2$
4	3.4261	0.5738	0.3292
6	5.8764	0.1235	0.0152
7	7.2252	-0.2252	0.0507
9	8.9331	0.0668	0.0044
10	10.8884	-0.8884	0.7893
12	11.9898	0.0101	0.0001
14	14.1926	-0.1926	0.0371
16	16.3954	-0.3954	0.1563
18	17.4968	0.5031	0.2532
20	19.5758	0.4241	0.1798
		0	1.8157

ويكون:

$$\hat{\sigma}_\varepsilon^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k-1} = \frac{\sum_{i=1}^{10}\hat{\varepsilon}_i^2}{n-k-1} = \frac{1.8157}{10-2-1} = 0.2593$$

نقوم بتحديد مصفوفة التباين–التباين المشترك للمقدرات $\hat{\Omega}_{\hat{\beta}}$:

$$\hat{\Omega}_{\hat{\beta}} = \hat{\sigma}_\varepsilon^2 (X'X)^{-1}$$

$$\hat{\Omega}_{\hat{\beta}} = 0.2593 \times \begin{pmatrix} 1.1725 & 0.0852 & -0.1921 \\ 0.0852 & 0.0284 & -0.0362 \\ -0.1921 & -0.0362 & 0.0547 \end{pmatrix}$$

نجد تباين كل مقدر بضرب تباين البواقي بكل عنصر من عناصر قطر المصفوفة $(X'X)^{-1}$:

$$\hat{\sigma}_{\hat{\beta}_0}^2 = 0.2593 \times 1.1725 = 0.3040 \rightarrow \hat{\sigma}_{\hat{\beta}_0} = 0.5513$$

$$\hat{\sigma}_{\hat{\beta}_1}^2 = 0.2593 \times 0.0284 = 0.0073 \rightarrow \hat{\sigma}_{\hat{\beta}_1} = 0.0858$$

$$\hat{\sigma}_{\hat{\beta}_2}^2 = 0.2593 \times 0.0547 = 0.0141 \rightarrow \hat{\sigma}_{\hat{\beta}_2} = 0.1190$$