تحليل الانحدار الخطي المتعدد

Multiple Linear Regression Analysis

المحاضـــــــرة الثانية

II-3. تقدير معالم نموذج الانحدار الخطي المتعدد

تستخدم طريقة المربعات الصغرى الاعتيادية لتقدير معالم نموذج الانحدار الخطي المتعدد.
حيث يأخذ نموذج الانحدار المقدر من العينة المقابل لمعادلة انحدار المجتمع الصيغة التالية:

$$Y_i = b_0 + b_1X_{i1} + b_2X_{i2} + \cdots + b_kX_{ik} + e_i \qquad (05)$$

$$\text{For } i = 1,2,\ldots\ldots\ldots,n$$

حيث أن:

$(b_0, b_1, b_2 \ldots, b_k)$: هي القيم المقدرة للمعالم $(\beta_0, \beta_1, \beta_2 \ldots, \beta_k)$.

(e_i): وهي تمثل البواقي أو الأخطاء، فالخطأ هو الفرق بين القيمة الفعلية للمتغير التابع والقيمة المقدرة لها.

(n): عدد المشاهدات أو حجم العينة.

يمكن كتابة المعادلة الخامسة في شكل مصفوفات كما يلي:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{12} & . & . & . & X_{1K} \\ 1 & X_{21} & X_{22} & . & . & . & X_{2K} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 1 & X_{n1} & X_{n2} & . & . & . & X_{nK} \end{pmatrix} \times \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ . \\ . \\ b_K \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{pmatrix} \qquad (06)$$

وباستخدام رموز المصفوفات يمكن اختصار كتابة نموذج الانحدار بالشكل التالي:

$$Y = Xb + e \qquad (07)$$

وبالرجوع دائما الى طريقة المربعات الصغرى الاعتيادية لتقدير معالم نموذج الانحدار الخطي المتعدد التي نحصل عليها بتدنية مجموع مربعات الأخطاء الى أصغر قيمة له. أي يجب أن تكون الدالة التالية نهاية صغرى:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{i1} - b_2 X_{i2} - \cdots - b_k X_{ik})^2$$

يمكن إعادة كتابة هذه المعادلة في شكل مصفوفات كما يلي:

$$\mathbf{e^T e} = (e_1 \;\; e_2 \;\; e_3 \; \ldots \; e_n) \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ . \\ . \\ . \\ e_n \end{pmatrix} = e_1^2 + e_2^2 + e_3^2 + \ldots + e_n^2 = \sum_{i=1}^{n} e_i^2$$

وبما أن:

$$Y = Xb + e \Rightarrow \mathbf{e} = Y - Xb$$

فإن:

$$e^T e = (Y - Xb)^T (Y - Xb)$$
$$e^T e = Y^T Y - 2b^T X^T Y + b^T X^T X b$$

وبتفاضل هذه المعادلة بالنسبة الى (b) ومساواة ناتج التفاضل بالصفر نحصل على:

$$\frac{\partial e^T e}{\partial b} = -2X^T Y + 2X^T X b = 0$$

$$\Rightarrow X^T X b = X^T Y$$

وبضرب طرفي هذه المعادلة بـ $(X^T X)^{-1}$ نتحصل على:

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

وبالتالي نجد:

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \qquad (08)$$

وبهذا نحصل على مقدرات معالم النموذج شريطة أن تكون رتبة المصفوفة $(X^T X)$ كاملة، أي أن تكون غير مفردة وذلك لإيجاد محددتها ومن ثم معكوسها $[(\mathbf{X}^T\mathbf{X})^{-1}]$.

مثال (01):

الجدول الموالي يوضح بيانات أوزان 50 طفلا تم اختيارهم عشوائيا من سجلات احدى مستشفيات الوطن، حيث وزن الطفل يمثل المتغير التابع (كلغ)، X_1 يمثل عمر الطفل (سنة)، و X_2 يمثل طول الطفل (سنتمتر). والمطلوب هو بناء نموذج الانحدار الخطي المتعدد (أي أوجد نموذج انحدار Y على X_1 و X_2).

جدول (2-1) بيانات أوزان وأعمار وأطوال 50 طفلا

الرقم	Y_i	X_1	X_2	الرقم	Y_i	X_1	X_2
1	11.5	3	84	26	5.3	0.33	57
2	16	5	95	27	6.5	0.75	63
3	6.5	0.5	65	28	13.5	3.83	92
4	17	4	100	29	4.5	0.25	53
5	8.5	1.33	70	30	15.5	4.75	98
6	8.8	1	70	31	16.5	4.67	102
7	22	6.17	118	32	11	1.75	80
8	13	3.42	95	33	17.5	5.25	96
9	12.5	3.67	94	34	14.55	4.83	103
10	15.5	5.42	97	35	10	2	83

11	9.5	1.17	76	36	4	0.17	52
12	15.5	4.42	96	37	3.5	0.08	50
13	9.5	1.17	73	38	8	1	70
14	14.5	2.75	100	39	8	1.33	72
15	19	6.25	115	40	14	3.75	95
16	9	1.5	76	41	1.75	0.17	31
17	14	4.25	98	42	3.2	0.08	46
18	10.5	2	80	43	5.55	0.33	46
19	6	0.42	63	44	2.75	0.08	51
20	15	5.58	105	45	1.35	0.01	46
21	13	3.42	94	46	5.5	0.58	36
22	21	6.17	118	47	4.5	0.08	46
23	12	3	90	48	3.25	0.02	35
24	17.5	5.25	100	49	3.3	00	49
25	5.5	0.33	56	50	1.4	0.08	40

الحل:

نجد أن المطلوب هو تقدير النموذج بواسطة العلاقة المقدرة التالية:

$$\hat{y}_i = b_0 + b_1X_{i1} + b_2X_{i2}$$

يتم تحديد مقدرات هذا النموذج باستخدام الصيغة التالية: $\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$

في البداية نقوم بوضع البيانات السابقة في قالب المصفوفات كما يلي:

$$Y = \begin{pmatrix} 11.5 \\ 16 \\ . \\ . \\ . \\ 1.4 \end{pmatrix} \quad , \quad X = \begin{pmatrix} 1 & 3 & 84 \\ 1 & 5 & 95 \\ . & . & . \\ . & . & . \\ . & . & . \\ 1 & 0.08 & 40 \end{pmatrix}$$

الخطوة الموالية هي حساب المصفوفة (X^TX) وذلك كما يلي:

$$(X^TX)=\begin{pmatrix}1 & 1 & . & . & . & 1\\ 3 & 5 & . & . & . & 0.08\\ 84 & 95 & . & . & . & 40\end{pmatrix}.\begin{pmatrix}1 & 3 & 84\\ 1 & 5 & 95\\ . & . & .\\ . & . & .\\ . & . & .\\ 1 & 0.08 & 40\end{pmatrix}$$

$$(X^TX)=\begin{pmatrix}50 & 117.4 & 3820\\ 117.4 & 491.7 & 11277.5\\ 3820 & 11277.5 & 320090\end{pmatrix}$$

بعد ذلك نقوم بحساب معكوس المصفوفة (X^TX) فنجدها تساوي:

$$(\mathbf{X^TX})^{-1}=\begin{pmatrix}0.789948 & 0.144239 & -0.014509\\ 0.144239 & 0.036936 & -0.003023\\ -0.014509 & -0.003023 & 0.000283\end{pmatrix}$$

وحساب المتجه (X^TY)**:**

$$(X^TY)=\begin{pmatrix}1 & 1 & . & . & . & 1\\ 3 & 5 & . & . & . & 0.08\\ 84 & 95 & . & . & . & 40\end{pmatrix}.\begin{pmatrix}11.5\\ 16\\ .\\ .\\ .\\ 1.4\end{pmatrix}=\begin{pmatrix}507.7\\ 1739.2\\ 45081.8\end{pmatrix}$$

بضرب $(X^TX)^{-1}$ ***في*** (X^TY) نحصل على تقديرات المربعات الضغرى التالية:

$$b=\begin{pmatrix}b_0\\ b_1\\ b_2\end{pmatrix}=(X^TX)^{-1}X^TY=$$

$$\begin{pmatrix}0.789948 & 0.144239 & -0.014509\\ 0.144239 & 0.036936 & -0.003023\\ -0.014509 & -0.003023 & 0.000283\end{pmatrix}.\begin{pmatrix}507.7\\ 1739.2\\ 45081.8\end{pmatrix}$$

$$b = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} -2.18190 \\ 1.20080 \\ 0.12457 \end{pmatrix}$$

وعليه فان نموذج انحدار وزن الطفل على عمره وطوله يأخذ الشكل التالي:

$$\hat{y}_i = -2.18190 + 1.20080X_1 + 0.12457X_2$$

يمكننا تفسير المعامل الثابت على أنه يمثل القيمة المقدرة للمتغير التابع عندما تكون قيم المتغيرات المفسرة مساوية للصفر. وفي الواقع نجد أن هذا التفسير ليس صحيحا في كل الحالات، فبأتباع هذا التفسير نجد أن الوزن المقدر يكون سالبا (2.1819 – كغ) عندما يكون عمر الطفل وطوله يساويان الصفر! . ولكن هل يوجد طفل عمره وطول صفر ؟ ! .

كما يجب ملاحظة أن مشاهدات العينة لا تحتوي على قيم صفرية لكل من متغيري العمر والطول.

أما معامل الانحدار الجزئي (1.20080) فيشير الى أن زيادة في عمر الطفل سنة واحدة تصحبها زيادة في وزنه بمقدار (1.20080كغ) بافتراض ثبات الطول. كذلك نجد أن الزيادة في طول الطفل بواحد سنتمتر تؤدي الى زيادة في وزنه بمقدار (0.12457 كغ) بافتراض ثبات العمر.

ملاحظة:

ان نموذج الانحدار الخطي المتعدد الموضح في العلاقة (01) يمكن كتابته بشكل مختصر كما يلي:

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \cdots + \beta_kX_K + \varepsilon \qquad (09)$$

وبالتالي فان نموذج الانحدار الخطي المتعدد المقدر يأخذ الشكل التالي:

$$\hat{y} = b_0 + b_1X_1 + b_2X_2 + \cdots + b_kX_K \qquad (10)$$