تحليل الانحدار الخطي المتعدد

Multiple Linear Regression Analysis

المحاضـــــــرة الأولى

II-1. نموذج الانحدار الخطي المتعدد

يوضح هذا النموذج العلاقة بين متغير تابع (Y) وأكثر من متغير مستقل. وبفرض أن (K) تمثل عدد المتغيرات المستقلة، فان نموذج الانحدار الخطي المتعدد يمكن التعبير عنه كامتداد لنموذج الانحدار الخطي البسيط وذلك كما يلي:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i \qquad (01)$$

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

يحتوي نموذج الانحدار الخطي المتعدد على مكونين: الأول يمثل **العنصر المحدد** ونعبر عنه بالمعادلة التالية: $(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik})$، والثاني يُمثل العنصر العشوائي (ε).

يُعرف المكون $(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik})$ **بنموذج انحدار المجتمع** ، حيث يُفترض أنه يمثل متوسط (Y) بشرط معرفة قيم محددة للمتغيرات المستقلة $(X_1, X_2, \ldots, X_k)$،أما (ε) فهو يمثل **الخطأ العشوائي** المرتبط بأي عنصر أو متغير في مجتمع الدراسة، وهو يمثل الاختلاف غي المفسر بين قيم (Y) عند مجموعة قيم المتغيرات المستقلة.

يمكن قياس الخطأ العشوائي بواسطة **تباين الخطأ** (σ_ε^2) وهو مثل المعالم $(\beta_0, \beta_1, \beta_2 \ldots, \beta_k)$ تكون غير معلومة ويجب تقديرها اعتمادا على بيانات عينة الدراسة.

وتظهر الأهمية القصوى للمقدار (σ_ε^2) بقياسه دقة توفيق المنحنى، وبالطبع فان التقدير الأقل لقيمة (σ_ε^2) يعني توفيق أفضل للبيانات.

يمكن كتابة المعادلة السابقة في جملة من المعادلات وذلك كما يلي:

$$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_k X_{1k} + \varepsilon_1$$

$$Y_2 = \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \cdots + \beta_k X_{2k} + \varepsilon_2$$

... ...

... ...

$$Y_N = \beta_0 + \beta_1 X_{N1} + \beta_2 X_{N2} + \cdots + \beta_k X_{Nk} + \varepsilon_N$$

وباستخدام صيغ المصفوفات يمكن كتابة هذه المعادلات على النحو التالي:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_N \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{12} & . & . & . & X_{1K} \\ 1 & X_{21} & X_{22} & . & . & . & X_{2K} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 1 & X_{N1} & X_{N2} & . & . & . & X_{NK} \end{pmatrix} \times \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_K \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_N \end{pmatrix} \qquad (02)$$

يمكن كتابة هذه المعادلة باستخدام رموز المصفوفات كما يلي:

$$Y = X\beta + \boldsymbol{\varepsilon} \qquad (03)$$

حيث:

(Y): متجه عمودي من الدرجة $(NX1)$ مشاهدة للمتغير التابع.

(X): مصفوفة البيانات (Data Matrix) من الدرجة [NX(k+1)] تحتوي على مشاهدات المتغيرات المستقلة، حيث يحتوي العمود الأول على قيم الواحد الصحيح لتمثيل المعامل الثابت.

(β): متجه أو شعاع عمودي من الدرجة [(k+1) X1] يحتوي على معالم نموذج الانحدار الخطي المتعدد المجهولة $(\beta_0, \beta_1, \beta_2 \dots, \beta_k)$ المراد تقديرها.

(ε): متجه أو شعاع عمودي من الدرجة (NX1) يحتوي على قيم المتغير العشوائي (ε_i).

II-2. فروض نموذج الانحدار الخطي المتعدد

ان فروض نموذج الانحدار الخطي المتعدد هي تعميم لفروض نموذج الانحدار الخطي البسيط التي تمت مناقشتها في الفصل السابق، والفرض الإضافي الوحيد المطلوب لنموذج الانحدار المتعدد هو عدم وجود علاقة ارتباط خطي بين متغيرين أو أكثر من المتغيرات المستقلة. وفيما يلي نستعرض هذه الفروض:

1-عدم وجود أخطاء تحديد ويشمل:

-وجود علاقة خطية بين المتغير التابع والمتغيرات المستقلة.
-أن يتضمن نموذج الانحدار الخطي المتغيرات المستقلة التي تساهم في تفسير المتغير التابع.

2-أن تكون قياسات المتغير التابع والمتغيرات المستقلة دقيقة وصحيحة.

3-أن يكون تباين أي متغير مستقل أكبر من الصفر.

4-القيمة المتوقعة لحد الخطأ العشوائي تساوي الصفر [E(ε) =0] أي:

$$E\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_N \end{pmatrix} = \begin{pmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ . \\ . \\ . \\ E(\varepsilon_N) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix} \qquad (04)$$

5-ثبات تباين حد الخطأ واستقلال قيم حدود الخطأ عن بعضها البعض.
6-أن يكون عدد المشاهدات أكبر من عدد المعالم المراد تقديرها أي: [n > (k+1)].
7-عدم وجود ارتباط خطي تام بين المتغيرات المستقلة.