

إن معادلة الانحدار الخطي المتعدد

(multiple linear regression equation)

هي التي تضم عدة متغيرات مستقلة وتكتب على الصورة التالية :

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + u_i \quad , \quad i=1,2,\dots,n$$

مع العلم أن u_i هو عبارة عن الفرق بين القيمة المشاهدة والقيمة المتوقعة للمتغير y_i ولذلك سمي بالخطأ أو الباقي أي أن

$$u_i = \hat{y}_i - y_i$$

من المعلوم أن نموذج الانحدار الخطي يمكن كتابته باستخدام المصفوفات على الشكل التالي :-

$$Y = \beta X + U$$

حيث أن U تشير إلى متجه الخطأ العشوائي و β تشير إلى متجه معالم الانحدار.

فروض نموذج الانحدار الخطي المتعدد

إن فروض الانحدار الخطي المتعدد هي تعميم لفروض نموذج الانحدار الخطي البسيط وفيما يلي نستعرض هذه الفروض:

١- أن تكون X عبارة عن مصفوفة متغيرات غير عشوائية

٢- أن يكون U متجة عشوائي وقيمته المتوقعة صفر أي أن

$$E(U) = 0$$

٣- أن يكون U غير مرتبطة ولها تباين σ^2 بمعنى أن

$$\text{Cov}(u) = E(uu') = \sigma^2 I$$

٤- u_i تتبع توزيع طبيعي مشترك
(jointly normal distribution) .

٥- عدم وجود أخطاء تحديد ويشمل:

• وجود علاقة خطية بين المتغير التابع y_i والمتغيرات المستقلة
 $x_{1i}, x_{2i}, \dots, x_{pi}$

• أن يتضمن نموذج الانحدار الخطي المتغيرات المستقلة التي تسهم في تفسير المتغير التابع، أي عند بناء النموذج يجب أن لا ندخل متغيرات ليس لها تأثير على المتغير وان لا نجهل في نفس الوقت إدخال متغيرات ذات تأثير عليه .

٦- أن تكون رتبة مصفوفة البيانات X كاملة ((Full rank أو مساوية لـ $(p+1)$ (عدد الأعمدة في المصفوفة). وهذا يعني أن تكون أعمدة المصفوفة X مستقلة خطياً، أي لا يكون هناك عمود من المصفوفة يمكن كتابته كتركيب خطي (Linear combination من الأعمدة الأخرى). وتعرف هذه الفرضية بفرضية عدم وجود ارتباط خطي تام بين المتغيرات المستقلة (No perfect multicollinearity) .

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

٨- أن يكون تباين أي متغير مستقل اكبر من الصفر أي :

$$\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 > 0 \quad \text{for } j=1,2,\dots,p$$

حيث إن x_{ij} المتغير المستقل رقم j و \bar{x}_j هو الوسط الحسابي له و n حجم العينة و p عدد المتغيرات المستقلة .