النمـوذج الخطـي العـام General Linear Model : يكـون المتغيـر المعتمـد Y دالـة خطيـة بدلالـة (k) مـن المتغيرات المستقلة: $(X_1,X_2,\ldots,X_k)$ وبذلك يصاغ نموذج الانحدار الخطي العام وفق العلاقة:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i$$

حيث ان:

$\beta_j \ , j = 1,\ldots,k$: تمثل معالم النموذج الجزئية التي تقيس استجابة المتغير التابع للمتغير المستقل المعني مع جعل بقيـة المتغيـرات المسـتقلة ثابتـة. وهـي تمثـل مقـدار التغيـر فـي Y للتغيـر بوحـدة واحـدة مـن X_i مـع ثبـات بقيـة المتغيرات المستقلة.

$u_i \ , i = 1,\ldots,n$: تمثل قيم المتغير العشوائي المجهولة.

والنموذج العام يكون:

$$Y = X\beta + u$$

حيث ان:

Y: متجه عمودي لمشاهدات المتغير المعتمد وبترتيب $(n \times 1)$.

X: مصفوفة تحتوي مشاهدات المتغيرات المستقلة X_1 ، X_2 ، . . . . ، X_k وعمودها الاول متغير وهمي للدلالة على المقطـع الصــادي. وترتيــب المصــفوفة $n \times (k+1)$. حيــث ان الصــف الأول مــن هــذه المصــفوفة هــو:

$(1 \quad X_{11} \quad X_{21} \ . \ . \ . \ X_{k1})$. وتسمى المصفوفة بمصفوفة المعلومات (data matrix).

β : متجه عمودي يحتوي معالم النموذج الخطي المراد تقديرها. وبترتيب $(k+1) \times 1$.

u : متجه عمودي يحوي قيم المتغير العشوائي المجهولة. وبترتيب $(n \times 1)$.

الطرف الأيمن في العلاقة يمثل الجزء المحدد بالمتغيرات المستقلة (Xβ) مضاف اليها الجزء الاحتمالي العشوائي (u) .

فرضيات نمـوذج الانحـدار الخطـي بصيغة المصـفوفات Assumptions of classical linear regression model in matrix form

- الفرضية الأولى: متوسط متجه المتغير العشوائي لكل عنصر من عناصره، يساوي صفراً.

أي:

$$E(u_i) = 0 \quad , \ i = 1,\ldots,n$$

وبالصيغة المصفوفية:

$$E(u) = E\begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ . \\ . \\ . \\ u_n \end{pmatrix} = \begin{bmatrix} E(u_1) \\ E(u_2) \\ E(u_3) \\ . \\ . \\ . \\ E(u_n) \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix}$$

• الفرضية الثانية: تباين المتغير العشوائي متجانس، قيمة التباين متساوية لكل مشاهدة. أي: (فرضية تجانس التباين Homoscedasticity)

$$E(u_i u_j) = \text{var}(u_i) = \sigma^2 \qquad , \qquad i = j$$

• الفرضـية الثالثـة: التبـاين المشـترك لأي مشـاهدتين مـن مشـاهدات المتغيـر العشـوائي، تسـاوي صـفراً أي:

$$E(u_i u_j) = 0 \qquad , \qquad i \neq j$$

وتسمى فرضية عدم وجود ارتباط ذاتي (No serial correlation)

وبالصيغة المصفوفية يمكن دمج الفرضية الثانية والثالثة كآلاتي:

$$E(uu') = E\begin{pmatrix} u_1 \\ u_2 \\ . \\ . \\ . \\ u_n \end{pmatrix} \begin{pmatrix} u_1 & u_2 & . & . & . & u_n \end{pmatrix}$$

$$E(uu') = E\begin{pmatrix} u_1^2 & u_1u_2 & u_1u_3 & . & . & . & u_1u_n \\ u_2u_1 & u_2^2 & u_2u_3 & . & . & . & u_2u_n \\ . & . & . & & & & . \\ . & . & . & & & & . \\ . & . & . & & & & . \\ u_nu_1 & u_nu_2 & u_nu_3 & . & . & & u_n^2 \end{pmatrix}$$

$$
= \begin{pmatrix}
Eu_1^2 & Eu_1u_2 & Eu_1u_3 & . & . & . & Eu_1u_n \\
Eu_2u_1 & Eu_2^2 & Eu_2u_3 & . & . & . & Eu_2u_n \\
. & . & . & & & . & \\
. & . & . & & & . & \\
. & . & . & & & . & \\
Eu_nu_1 & Eu_nu_2 & Eu_nu_3 & . & . & & Eu_n^2
\end{pmatrix}
$$

$$
= \begin{pmatrix}
\sigma^2 & 0 & \cdots & 0 \\
0 & \sigma^2 & \cdots & 0 \\
\vdots & \cdots & \ddots & \vdots \\
0 & 0 & \cdots & \sigma^2
\end{pmatrix} = \sigma^2 \begin{pmatrix}
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \cdots & \ddots & \vdots \\
0 & 0 & \cdots & 1
\end{pmatrix}
$$

$$E(uu') = \sigma^2 I_n$$

وهذه المصفوفة تمثل مصفوفة التباين. والتباين المشترك للمتغير العشوائي u بالصيغة المصفوفية، عناصر القطر الرئيسي (main diagonal) تمثل تباين مشاهدات المتغير العشوائي. بينما عناصر المثلث العلوي تساوي عناصر المثلث السفلي لأنها تمثل التباين المشترك لمشاهدات المتغير العشوائي. وغني عن الذكر المصفوفة تكون متماثلة .(symmetric)

اما مصفوفة (I_n) هي مصفوفة احادية او متطابقة (Unit or Identity Matrix) وهي ذات رتبة $(n \times n)$

الفرضية الرابعة: المتغيرات $X_1, X_2, \ldots, X_k$ متغيرات ثابتة للعينة المختارة فهي غير عشوائية بمعنى ان المتغيرات التوضيحية تكون متغيرات خارجية (Exogenous) ، وبذلك فان المصفوفة X ذات الترتيب $n \times (k+1)$ غير عشوائية. بمعنى تحتوي على أرقام ثابتة.

الفرضية الخامسة: لا توجد علاقة خطية تامة بين المتغيرات التي تمثل المصفوفة X . بمعنى لا وجود للتعدد الخطي. ويمكن تجزئة الفرضية كالآتي:

$$r_{x_0 x_j} = 0 \quad \forall \quad j = 1,2,\ldots,k \qquad (١)$$

وتقرأ باستقلال العمود الأول (والذي مشاهداته واحد) مع أي من أعمدة المصفوفة الأخرى.

وعملياً يعني وجود تغييرات واضحة وملموسة في مشاهدات كل متغير من المتغيرات المستقلة (التوضيحية) X_1، . . . ، X_k .

$$r_{x_i x_j} = 0 \quad \forall i \neq j \quad , \quad i, j = 1,2,\dots,k \qquad (٢)$$

وبذلك يفترض استقلال المتغيرات التوضيحية عن بعضها الآخر. فالأعمدة الثاني والثالث ...... والعمود (k+1) مستقلة عن بعضها الآخر.

ويمكن عرض هذه الفرضية بالصيغة المصفوفية:

$$\rho(X) = k + 1 \; \langle \; n \qquad . \; . \; . \qquad (3)$$

$\rho(X)$: رتبة مصفوفة المعلومات (X) وتكون تامة من ناحية الأعمدة وهي أقل من عدد المشاهدات (n) ، اي ان عدد المشاهدات يجب ان يكون اكبر من عدد المعلمات المطلوب تقديرها .

الفرضية السادسة: يفترض ان مشاهدات المتغير العشوائي u تتوزع بشكل متماثل أي لها توزيع طبيعي (Normal) . وبصيغة المصفوفات:

ان متجه المتغير العشوائي u له توزيع طبيعي متعدد. $u \sim N$

وبدمج الفرضيات (2) و (3) و (6) يمكن صياغتها مصفوفياً كالآتي:

$$u \sim N(0, \sigma^2 I_n) \qquad . \; . \; . \qquad (4)$$