تحليل الانحدار Regression Analysis

الانحدار الخطي المتعدد

في الانحدار الخطي البسيط لدينا متغير مستقل (x_i) واحد وكذلك متغير معتمد واحد

Y_i	X_i
y_1	x_1
y_2	x_2
.	.
.	.
y_n	x_n

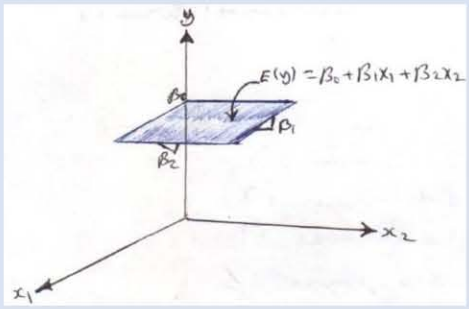
أما في حالة الانحدار الخطي المتعدد فيكون لدينا m من المتغيرات المستقلة كل منها يحتوي على n من المشاهدات وكما يلي

Y_i	X_{i1}	X_{i2}	..... x_{im}
y_1	x_{11}	x_{12}	..... x_{1m}
y_2	x_{21}	x_{22}	..... x_{2m}
.	.	.	.
.	.	.	.
y_n	x_{n1}	x_{n2}	..... x_{nm}

ويمكن تمثيل العلاقة بين المتغير المعتمد y_i والمتغيرات المستقلة x_i's، فلو كان لدينا مثلاً متغيرين مستقلين فقط x_{i1} وx_{i2} فمعادلة الانحدار المتعدد في هذه الحالة هي سيصبح متوسط قيم y لمختلف قيم x_1 وx_2

$$Y_i = \beta_0 + \beta_{i1} X_{i1} + \beta_{i2} Xi_2 + U_I$$

ويكون الشكل سطح بثلاث أبعاد كالأتي:



لذا فانه وبصورة عامة يمكن تمثيل العلاقة الدالية Functional Relationship بين المتغير المعتمد y والمتغيرات المستقلة x_i's في تحليل الانحدار المتعدد كما يلي:

$$Y_i = \beta_0 + \beta_{i1} X_{i1} + \beta_{i2} Xi_2 + \ldots + \beta_{iK} X_{iK} + \varepsilon_i$$

وتسمى هذه المعادلة بنموذج من الدرجة الأولى First order model

أو الانحدار السطحي surface model

أو الاستجابة السطحية Regression surface model

حيث ان :

$X_{i1}, X_{i2}, \ldots, X_{im}$ عبارة عن قيم ثابتة ل m من المتغيرات المستقلة وتسمى ثابته لان ليس لها توزيع عشوائي

$\beta_0, \beta_1, \beta_2 \ldots\ldots\ldots, \beta_m$:هي ثوابت او معلمات معادلة الانحدار.

β_0 :محل تقاطع مستوى الانحدار regression plane مع المحور y

أو متوسط الاستجابة عندما تكون كافة المتغيرات المستقلة تساوي صفر

β_j : يمثل معامل الانحدار الجزئي ل y على x_j عندما تكون بقية المتغيرات المستقلة ثابتة.

أو مقدار التغير في y عند زيادة وحدة واحدة في x_j عندما تكون بقية المتغيرات ثابتة.

ε_i الخطأ العشوائي

وباستخدام المصفوفات يمكن كتابة النموذج أعلاه كما يلي

حيث أن :

$$\underline{y} = X\underline{\beta} + \underline{u}$$

حيث أن :

$\underline{Y}$: تمثل متجه لمشاهدات متغير الاستجابة

X : تمثل مصفوفة لمشاهدات المتغيرات التوضيحية

$\underline{\beta}$: متجه معلمات الانحدار .

$\underline{e}$:متجه الخطأ العشوائي

فروض التحليل Assumption of analysis

1- إن Y متغير عشوائي وقيمته مستقلة إحصائيا الواحد عن الأخر ويتوزع توزيعا طبيعيا بوسط حسابي قدره $\mu_{y|x_1x_2...x_m}$ وتباين $\sigma^2_{y|x_1x_2...x_m}$

أي أن

$$y \sim N(\mu_{y|x_1x_2...x_m}, \sigma^2_{y|x_1x_2...x_m})$$

$$\mu_{y|x_1x_2...x_m} = E(Y_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_m X_{im}$$

$$\sigma^2_{y|x_1x_2...x_m} = \sigma^2_{(\beta_0+\beta_1 X_1+\beta_2 X_2+...+\beta_m X_{im})} = \sigma^2_e = \sigma^2 I_n$$

2- إن ε_i هو خطأ عشوائي يتبع التوزيع الطبيعي بوسط حسابي $E(\varepsilon_i)=0$ وتباين ثابت مقداره σ^2_ε .

$$e_i \quad N(0, \sigma^2_\varepsilon)$$

3- إن قيمة الخطأ ε_i في فترة معينة غير مرتبطة بقيمتها في فترة أخرى ولذلك فان التباين المشترك بينهما يساوي صفراً أي أن

$$cov(\varepsilon_i \varepsilon_i') = 0$$

4- إن قيم متغير المستقلة X_j ثابتة fixed أي أن معطاة ولا تتوزع عشوائيا

5- عدم وجود علاقة خطية تامة او غير تامة بين المتغيرات المستقلة أي غياب مشكلة تعدد العلاقة الخطية Multicollinearity.

تقدير المعلمات بطريقة المربعات الصغرى

$$Y_1 = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{im} + \varepsilon_i$$

$$\varepsilon_i = Y_1 - (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{im})$$

$$\Rightarrow \varepsilon_i = Y_i - \beta_0 - \sum_{j=1}^{m} \beta_j X_{ij}$$

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \sum_{j=1}^{m} \hat{\beta}_j X_{ij})^2 = K$$

وبأخذ المشتقة الجزئية ل K ومساواتها بالصفر نحصل على:

$$\frac{\partial K}{\partial \beta_0} = -2 \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \sum_{j=1}^{m} \hat{\beta}_j X_{ij}) = 0$$

$$\frac{\partial K}{\partial \beta_j} = -2 \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \sum_{j=1}^{m} \hat{\beta}_j X_{ij}) X_{ij} = 0$$

وبحل المعادلتين أعلاه نحصل على المعادلات الطبيعية وكما يلي

$$\begin{bmatrix} n & \sum_{i=1}^{n} X_{i1} & . & . & \sum_{i=1}^{n} X_{im} \\ \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X^2_{i1} & . & . & \sum_{i=1}^{n} X_{i1} X_{im} \\ . & . & & & . \\ . & . & & & . \\ . & . & & & . \\ \sum_{i=1}^{n} X_{1k} & \sum_{i=1}^{n} X_{im} X_{i1} & . & . & \sum_{i=1}^{n} X^2_{ik} \end{bmatrix} * \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ . \\ . \\ . \\ . \\ \hat{\beta}_m \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{i1} y_i \\ \sum x_{i2} y_i \\ . \\ . \\ . \\ \sum x_{im} y_i \end{bmatrix}$$

$$(\underline{X'X}) \qquad \underline{\hat{\beta}} \qquad (\underline{X'Y})$$

$$\underline{X} = \begin{bmatrix} 1 & X_{11} & . & . & . & X_{1m} \\ 1 & X_{21} & . & . & . & X_{2m} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ 1 & X_{n1} & . & . & . & X_{nm} \end{bmatrix}_{n(m+1)}$$

هذا يعني

$$(X'X)\hat{\underline{\beta}} = (X'\underline{Y})$$

وبضرب كلا الطرفين في $(X'X)^{-1}$ فان مقدرات المربعات الصغرى الاعتيادية يعطى بالشكل التالي:

$$(X'X)^{-1}(X'X)\hat{\underline{\beta}} = (X'X)^{-1}(X'\underline{Y})$$

$$\Rightarrow \hat{\underline{\beta}}_{ols} = (X'X)^{-1}X'\underline{Y}$$

$$\hat{Y}_1 = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \ldots + \hat{\beta}_m X_{im}$$
$$\hat{Y}_2 = \hat{\beta}_0 + \hat{\beta}_1 X_{21} + \hat{\beta}_2 X_{22} + \ldots + \hat{\beta}_m X_{2m}$$

$$\Rightarrow \underline{\hat{y}} = \underline{x}\hat{\underline{\beta}}$$

$$\underline{e} = \underline{y} - \underline{\hat{y}} \quad \triangleright \quad \underline{e} = \underline{y} - \underline{x}\hat{\underline{\beta}}$$

$$(X'X) = \underline{C} = \begin{bmatrix} C_{00} & C_{01} & \ldots\ldots C_{0m} \\ c_{10} & C_{11} & \ldots\ldots C_{1m} \\ . & & \\ . & & \\ . & & \\ C_{m0} & C_{m1} & \ldots\ldots C_{mm} \end{bmatrix}_{(m+1)(m+1)} \qquad \triangleright V(\hat{\underline{\beta}}) = \hat{\sigma}^2 (X'X)^{-1} = \hat{\sigma}^2 \underline{C}$$