تحليل الانحـدار الخطي المتعدد
Multiple Regression Analysis

الغرض من التحليل

يهتم تحليل الانحدار الخطي المتعدد بدراسـة وتحليل أثر عدة متغيرات مستقلة كمية على متغير تابع كمي.

نموذج الانحـــدار الخطي المتعدد

بفرض أن المتغير y يعبر عن المتغير التابع، والمتغيرات $(x_1, x_2, ..., x_k)$ تعبر عن k من المتغيرات المستقلة، وأن عدد المشاهدات هي n، فإن المشاهدة التابعة y_i، $i = 1,2,...,n$ يمكن التعبير عنها كدالة خطية في مجموعة المشاهدات المفسرة $(x_{i1}, x_{i2}, ..., x_{ip})$ كما يلي:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \varepsilon_i \qquad \textbf{(7.1)}$$

حيث أن $(\beta_0, \beta_1, \beta_2, ..., \beta_k)$ تعبر عن معاملات الانحدار، ε_i يعبر عن الخطأ العشوائي للمشاهدة رقم i، $i = 1,2,...,n$. وحيث أن عدد المشاهدات هي n، يكون لدينا n من المعادلات يمكن صياغتها في صورة مصفوفات كما يلي:

$$\begin{bmatrix} y_1 \\ y_2 \\ .... \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} ... x_{1k} \\ 1 & x_{21} & x_{22} ... x_{2k} \\ & & ..... \\ 1 & x_{n1} & x_{n2} ... x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ ..... \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ .... \\ \varepsilon_n \end{bmatrix} \qquad \textbf{(7.2)}$$

$$\mathrm{Y} = \mathrm{XB} + \mathrm{E}$$

حيث أن:

Y : يعبر عن متجه المشاهدات التابعة، وهو من درجة $(n \times 1)$، والعنصر رقم i في هذا المتجه هو y_i.

X : تمثل مصفوفة المشاهدات المستقلة (المفسرة)، وهي من درجة $(n \times (k+1))$، والصف رقم i في هذه المصفوفة هو: $(1 \;\; x_{i1} \;\; x_{i2} ... x_{ik})$.

B : يعبر عن متجه معاملات الانحدار، وهو من الدرجة $((k+1) \times 1)$.

E : يعبر عن متجه الأخطاء العشوائية، وهو من درجة $(n \times 1)$، والعنصر رقم i هو الخطأ العشوائي ε_i.

افتراضات النموذج

يستند نموذج الانحدار المتعدد والمبين بالمعادلة (7.2) على عدة افتراضات هي:

١- مصفوفة المتغيرات المستقلة X محددة، ومعطاة Fixed، فهي مقاسه بدون أخطاء.

٢- المتغيرات المستقلة $(x_{i1} \;\; x_{i2} ... x_{ik})$ مستقلة إحصائيا، ويعني ذلك وجود استقلال خطي بين أعمدة

$(k+1)$ X X

$$(Rank(\mathrm{X}) = k+1) < n \qquad \textbf{(7.3)}$$

٣- يوجد استقلال إحصائي بين المشاهدات المستقلة $(x_{i1}\ x_{i2}...x_{ik})$ **، والخطأ العشوائي** ε_i **، أي أن أعمدة المصفوفة** X **مستقلة خطيا عن متجه الأخطاء العشوائية** E **، ويعبر عن ذلك رياضيا كما يلي:**

$$Cov(\mathrm{X},\mathrm{E}) = E(\mathrm{X}'\mathrm{E}) - [E(\mathrm{X})]'[E(\mathrm{E})] = 0 \qquad \textbf{(7.4)}$$

٤- الخطاء العشوائي ε_i **،** $i = 1,2,...,n$ **له توزيع طبيعي متوسطه صفرا، وتباين** σ^2 **، ثابت من مشاهدة إلى أخرى، أي أن** $\varepsilon_i \sim N(0,\sigma^2)$ **، كما يفترض أن الأخطاء** $(\varepsilon_1,\varepsilon_2,...,\varepsilon_n)$ **مستقلة إحصائيا، ويعبر عن ذلك رياضيا كما يلي:**

$$E(\varepsilon_i) = 0$$

$$Cov(\varepsilon_i,\varepsilon_j) = E(\varepsilon_i\varepsilon_j) = \begin{cases} \sigma^2 \ if\ i = j \\ 0 \ if\ i \neq j \end{cases} \qquad \textbf{(7.5)}$$

أي أن متجه الأخطاء E **يتبع توزيع طبيعي متعدد متوسطه صفرا وله مصفوفة تباين** Σ **، أي أن**

$$\mathrm{E} \sim N_n(0,\Sigma) \qquad \textbf{(7.6)}$$

حيث أن المصفوفة Σ **مصفوفة متماثلة ومن الدرجة** $(n \times n)$ **، ويعبر عنها كما يلي:**

$$\Sigma = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ & & & \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & & & \\ 0 & 0 & \cdots & 1 \end{bmatrix} \qquad \textbf{(7.6)}$$

تقديرات المربعات الصغرى لمعالم النموذج

يحتوي النموذج (7.2) على $(k+1)$ **من المعاملات هي المتجه** $\mathrm{B} = (\beta_0,\beta_1,...,\beta_k)'$ **، يمكن تقديرها بطريقة المربعات الصغرى (OLS)، وهذا التقدير هو:**

$$\hat{\mathrm{B}} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{Y} \qquad \textbf{(7.7)}$$

وتحت تحقق الافتراضات أعلاه، يكون التقدير $\hat{\mathrm{B}}$ **هو التقدير الخطي الأفضل غير المتحيز (BLUE).**

تطبيق (٧-٨) صـ ٣٣٤

y_i **: تعبر عن درجة تفضيل صنف،** x_{i1} **: الرطوبة،** x_{i2} **: حلاوة المنتج.**

- **النموذج الخطي المقترح:** $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \ ,\ i = 1,2,...,16$
- **تقديرات المربعات الصغرى:**

$$\hat{\mathrm{B}} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{Y}$$

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{bmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y \\ \sum x_1 y \\ \sum x_2 y \end{bmatrix}$$

$$(X'X) = \begin{bmatrix} 16 & 116 & 48 \\ 116 & 928 & 352 \\ 48 & 352 & 160 \end{bmatrix} \quad X'Y = \begin{bmatrix} 1308 \\ 9862 \\ 3994 \end{bmatrix}$$

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	41.302	3.482		11.862	.000	33.780	48.824					
	X1	4.203	.355	.884	11.829	.000	3.436	4.971	.916	.957	.879	.989	1.012
	X2	3.324	.829	.300	4.011	.001	1.534	5.114	.395	.744	.298	.989	1.012

a Dependent Variable: Y

$$\hat{\mathrm{B}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 41.302 \\ 4.203 \\ 3.324 \end{pmatrix}$$

إذا معادلة الانحدار هي:

$$\hat{y}_i = 41.302 + 4.203x_{i1} + 3.324x_{i2}$$

بعض مؤشرات جودة النموذج

من مؤشرات جودة النموذج معامل التحديد R^2 ، وفي المحاضرة السابقة وضحنا أن مجموع المربعات الكلي SST يمكن تقسيمه إلى جزئين كما هو مبين بالمعادلة التالية:

$$SST = SSR + SSE$$

ويكون معامل التحديد هو:

$$R^2 = \frac{SSR}{SST}$$

وفي حالة استخدام المصفوفات يلاحظ أن هذه المجاميع تحسب كالتالي:

$$SST = \sum (y - \bar{y})^2 = \mathrm{Y}'Y - \frac{(\sum y)^2}{n} \qquad \textbf{(7.8)}$$

$$SSR = \sum (\hat{y} - \bar{y})^2 = \hat{\mathrm{B}}'\mathrm{X}'Y - \frac{(\sum y)^2}{n}$$

$$SSE = \sum (y - \hat{y})^2 = SST - SSR = \mathrm{Y}'\mathrm{Y} - \hat{\mathrm{B}}'\mathrm{X}'Y$$

وفي التمرين السابق نجد أن معامل التحديد قيمته هي: $R^2 = 0.928$ **، وتدل على أن الرطوبة** x_1 **، وحلاوة المنتج** x_2 **يفسران 92.8% من الاختلافات في التفضيل** y **، وأن النسبة المتبقية 7.2% ترجع للأخطاء العشوائية، ومن المتوقع أن يكون هذا النموذج توفيق جيد للعلاقة بين متغير درجة التفضيل كمتغي تابع، والرطوبة وحلاوة المنتج كمتغيرين مفسرين.**

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.963(a)	.928	.917	3.29556	.928	84.056	2	13	.000
a Predictors: (Constant), X2, X1									

اختبار صلاحية النموذج

أولا : اختبار جودة النموذج

١- صياغة الفرضين العدم H_o **، والبديل** H_A

$$H_o : \beta_1 = \beta_2 = ... = \beta_k = 0 \quad , \quad H_A : at\ least\ two\ unequal$$

النموذج مناسب **النموذج غير مناسب**

٢- إحصائية الاختبار

$$F^* = \frac{\left(\frac{SSR}{k}\right)}{\left(\frac{SSE}{(n-k-1)}\right)} = \frac{\left(\frac{R^2}{k}\right)}{\left(\frac{(1-R^2)}{(n-k-1)}\right)} \quad , \sim F_{(k,n-k-1)} \qquad \textit{(16)}$$

٣- تحديد مناطق الرفض والقبول

بالكشف فى جدول توزيع F **عند مستوى المعنوية المحدد** α **، ودرجات حرية البسط** (k) **، ودرجات حرية مقام** $(n-k-1)$ **، يمكن استخراج القيمة الحرجة Critical value** $F^{(1-\alpha)}_{(k,n-k-1)}$ **، وتحديد مناطق الرفض والقبول.**

٤- القرار : إذا وقعت F^* **المحسوبة فى منطقة الرفض ، فإنه لا يمكن قبول الفرض العدم** H_o **، ويستدل**

ذلك على أن النموذج ، مناسب فى تمثيل العلاقة الخطية المفترضة بين المتغير التابع والمتغيرات المفسرة.

وفي التطبيق السابق يمكن إجراء اختبار صلاحية النموذج كما يلي:

١- صياغة الفرضين العدم H_o **، والبديل** H_A

$$H_o : \beta_1 = \beta_2 = 0 \ , \quad H_A : at \ least \ two \ unequal$$

النموذج مناسب **النموذج غير مناسب**

٢- إحصائية الاختبار

$$F^* = \frac{\left(\frac{SSR}{k}\right)}{\left(\frac{SSE}{(n-k-1)}\right)} = \frac{\left(\frac{1825.811}{2}\right)}{\left(\frac{(141.189)}{(16-2-1)}\right)} = \frac{912.906}{10.861} = 84.056 \qquad \textbf{(16)}$$

٣- تحديد مناطق الرفض والقبول

بالكشف فى جدول توزيع F **عند مستوى المعنوية** $\alpha = 0.05$ **، ودرجات حرية البسط** $(k = 2)$ **، ودرجات حرية مقام** $(n-k-1 = 13)$ **، يمكن استخراج القيمة الحرجة وهي:** $F_{(2,13)}^{(0..95)} = 3.89$ **.**

القرار : بما أن قيمة F^* **تزيد عن قيمة** F **الجدولية ويستدل من ذلك على أن النموذج ، مناسب فى تمثيل العلاقة الخطية المفترضة بين المتغير التابع والمتغيرين المفسران.**

ANOVA(b)						
Model		**Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
1	**Regression**	1825.811	2	912.906	84.056	.000(a)
	Residual	141.189	13	10.861		
	Total	1967.000	15			
a Predictors: (Constant), X2, X1						
b Dependent Variable: Y						