تحليل الانحدار الخطي المتعدد

Multiple Linear regression Analysis

الغرض من استخدام التحليل:

دراسة وتحليل أثر عدة متغيرات مستقلة كمية على متغير تابع كمي.

نموذج الانحدار الخطي المتعدد:

يمكن تمثيل نموذج الانحدار الخطي المتعدد في الشكل الآتي:

$$\begin{aligned} y_i &= \mu_{y_i|x_{i1},x_{i2},\dots,x_{ip}} + \varepsilon_i \\ &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i \end{aligned} \qquad (1)$$

حيث أن:

y_i : المشاهدة للمتغير التابع على المفردة رقم i.

$x_{i1}, x_{i2},\dots, x_{ip}$: مجموعة المشاهدات للمتغيرات المفسرة أو المستقلة على المفردة رقم i ، وعددها p متغير مستقل.

$\beta_0, \beta_1, \beta_2,\dots, \beta_p$: تعبر عن معاملات الانحدار والتي تعكس آثار المتغيرات المفسرة.

ε_i : خطأ عشوائي.

وبفرض أن لدينا عينة عشوائية حجمها n ، مفردة، يصبح لدينا n من المعادلات على الصورة رقم (1) ويعبر عنها كالتالي:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_p x_{1p} + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_p x_{2p} + \varepsilon_2 \\ &\dots\dots\dots \quad \dots\dots\dots \quad \dots\dots\dots \quad \dots\dots\dots \quad \dots\dots\dots \quad .. \\ y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_p x_{np} + \varepsilon_n \end{aligned} \qquad (2)$$

.

ومن ثم يمكن وضع المعادلات الخطية (2) في صورة معادلة مصفوفية كما يلي:

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \multicolumn{5}{c}{\ldots\ldots\ldots \quad \ldots\ldots\ldots \ .} \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix} \qquad (2)$$

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

افتراضات النموذج:

1- المشاهدات التابعة $y_1, y_2, \ldots, y_n$ مشاهدات مستقلة من مجتمع له توزيع طبيعي متعدد متوسطه هو $\mu_{y|\underline{\mathbf{x}}} = (\mu_{y_1|\underline{\mathbf{x}}_1}, \mu_{y_2|\underline{\mathbf{x}}_2}, \ldots, \mu_{y_n|\underline{\mathbf{x}}_n})'$ ، وله مصفوفة تباين: $\Sigma_{\underline{\mathbf{y}}|\underline{\mathbf{x}}} = \sigma^2_{y|\underline{\mathbf{x}}}\ I(n)$ هي.

2- المشاهدات المفسرة $x_{i1}, x_{i2}, \ldots, x_{ip}$ ثابتة Fixed ومعطاه، كما أنها مستقلة إحصائيا، وهذا يعني أن رتبة المصفوفة $rank\ (\mathbf{x}) = (p+1) < n$.

3- الأخطاء العشوائية $[\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n]$ متغيرات عشوائية لها توزيع طبيعي متعدد متوسطه صفر وتباينه $\Sigma_{\underline{\mathbf{y}}|\underline{\mathbf{x}}} = \sigma^2_{y|\underline{\mathbf{x}}}\ I(n)$.

4- الأخطاء العشوائية مستقلة إحصائيا عن المشاهدات المفسرة.

تقدير المربعات الصغرى (OLS) لمعاملات الانحدار:

تقدير المربعات الصغرى لمتجه معاملات الانحدار $\boldsymbol{\beta} = [\beta_0, \beta_1, \ldots, \beta_p]'$ هو الذي يجعل مجموع مربعت الأخطاء العشوائية $SSE = \sum \varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$ أصغر ما يمكن، ويعبر عنه بالمعادلة التالية:

$$\hat{\boldsymbol{\beta}} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}$$

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \cdots \\ \hat{\beta}_p \end{pmatrix} = \begin{pmatrix} n & \sum x_1 & \cdots & \sum x_p \\ \sum x_1 & \sum x_1^2 & \cdots & \sum x_1 x_p \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \sum x_p & \sum x_1 x_p & \cdots & \sum x_p^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y \\ \sum x_1 y \\ \ldots \\ \sum x_p y \end{pmatrix} \qquad (3)$$

ويمكن استخدام برنامج Excel في إجراء العمليات الجبرية المتعلقة بالمصفوفات

تطبيق (1):

فيما يلي بيانات عن الإنفاق الاستهلاكي العائلي y بالألف ريال خلال الشهر، و الدخل العائلي x_1، وعدد أفراد الأسرة x_2 لـ 12 أسرة.

الإنفاق الاستهلاكي	الدخل العائلي	عدد أفراد الأسرة
y	x_1	x_2
2.5	**9.5**	**1**
2.8	**9.8**	**3**
3.8	**10.0**	**4**
4.5	**11.1**	**4**
4.7	**12.0**	**4**
4.7	**12.8**	**5**
4.8	**13.4**	**5**
5.2	**15.4**	**5**
6.4	**15.6**	**6**
7.1	**15.9**	**6**
8.1	**20.6**	**6**
8.9	**21.8**	**8**

والمطلوب

1- كتابة الشكل الرياضي للنموذج في المجتمع.
2- حساب تقدير المربعات الصغرى لمعادلة انحدار الإنفاق على الدخل وعدد أفراد الأسرة.
3- حساب معامل التحديد وعلام يدل؟
4- تكوين جدول تحليل التباين.
5- اختبار صلاحية النموذج الخطي.
6- تفسير معاملات النموذج واختبار معنوياتها.
7- تفسير فترات الثقة 95% للمعاملات.
8- متوسط الإنفاق المتنبأ به لأسرة دخلها 20 ألف وعدد أفرادها 7، وما هوا حدا الثقة لهذا المتوسط، $\alpha=0.05$.

حل التطبيق

1- كتابة الشكل الرياضي للنموذج في المجتمع.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

2- حساب تقدير المربعات الصغرى لمعادلة انحدار الإنفاق على الدخل وعدد أفراد الأسرة.

$$OLS \quad Estimation \quad : \quad \hat{\boldsymbol{\beta}} = (\mathbf{x}'\mathbf{x})^{-1}(\mathbf{x}'\mathbf{y})$$

$$(\mathbf{x}'\mathbf{x}) = \begin{pmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 \end{pmatrix} = \begin{pmatrix} 12 & 167.9 & 57 \\ 167.9 & 2529.83 & 866.3 \\ 57 & 866.3 & 305 \end{pmatrix}$$

$$(\mathbf{x}'\mathbf{x})^{-1} = \begin{pmatrix} 1.208 & -0.105 & 0.072 \\ -0.105 & 0.024 & -0.047 \\ 0.072 & -0.047 & 0.124 \end{pmatrix} \qquad (\mathbf{x}'\mathbf{y}) = \begin{pmatrix} \sum y \\ \sum x_1 y \\ \sum x_2 y \end{pmatrix} = \begin{pmatrix} 63.5 \\ 973.71 \\ 337.2 \end{pmatrix}$$

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 1.208 & -0.105 & 0.072 \\ -0.105 & 0.024 & -0.047 \\ 0.072 & -0.047 & 0.124 \end{pmatrix} \begin{pmatrix} 63.5 \\ 973.71 \\ 337.2 \end{pmatrix} = \begin{pmatrix} -1.088 \\ 0.325 \\ 0.387 \end{pmatrix}$$

ومن ثم تأخذ معادلة التنبؤ الصورة التالية.

$$\begin{aligned} \hat{y} &= \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \\ &= -1.088 + 0.325\ x_1 + 0.387\ x_2 \end{aligned}$$

3- حساب معامل التحديد وبيان مدلوله.

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

$$\begin{aligned} SSY &= \mathbf{y}'\mathbf{y} - CF = \sum y^2 - \frac{(\sum y)^2}{n} = 379.23 - \frac{(63.5)^2}{12} \\ &= 379.23 - 336.02 = 43.21 \end{aligned}$$

$$SSR = \hat{\boldsymbol{\beta}}'\mathbf{x}'\mathbf{y} - CF = \begin{pmatrix} -1.088 & 0.325 & 0.387 \end{pmatrix} \begin{pmatrix} 63.5 \\ 973.71 \\ 337.2 \end{pmatrix} - 336.02$$

$$= 377.45 - 336.02 = 41.43$$

$$R^2 = \frac{SSR}{SSY} = \frac{41.43}{43.21} = 0.959$$

الدخل وعدد أفراد الأسرة متغيران كمتغيران مستقلان يفسران 95.9% من الاختلافات في الإنفاق.

4- تكوين جدول تحليل التباين.

S.O.V	SS	df	MS= SS / df	F*
Reg.	SSR	p	MSR	MSR / MSE
Error	SSE	n-p-1	MSE	
Total	SSY	n-1		

$$SSE = SSY - SSR = 43.21 - 41.43 = 1.78$$

$$p = 2, \; n = 12$$

جدول تحليل التباين

S.O.V	SS	df	MS	F*	Sig
$Reg_{(X1,X2)}$	41.43	2	20.716	104.89	0.0000006
Error	1.78	9	0.198		
Total	43.21	11			

5- اختبار صلاحية النموذج الخطي عند مستوى معنوية 5%.

الفرض العدم: النموذج الخطي غير منسب للتنبؤ $H_0 : \beta_1 = \beta_2 = 0$

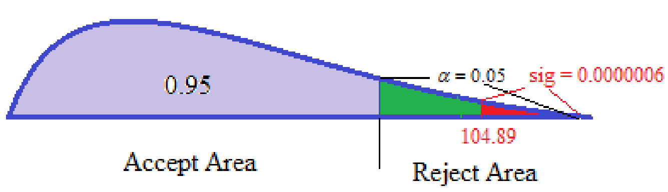
الفرض البديل: النموذج الخطي مناسب للتنبؤ $H_1 : \text{at least one of } \beta_j \neq 0, \; j = 1,2$

إحصائية الاختبار: $F^* = \dfrac{MSR}{MSE} = \dfrac{20.716}{0.198} = 104.89$

مستوى المعنوية : $\alpha = 0.05$

الاحتمال المشاهد (المعنوية المحسوبة) : $p \text{ value } (Sig.) = P(F_{(2,9)} > 104.89) = 0.0000006$

القرار: بما أن المعنوية المحسوبة $(Sig.) = 0.0000006$ أقل من مستوى المعنوية $\alpha = 0.05$ لذا نرفض الفرض العدم ونقبل الفرض البديل، ويستدل من ذلك على أن النموذج الخطي مناسب للتنبؤ، ويبين الشكل التالي منحنى التوزيع الاحتمالي F ومناطق الرفض والقبول.



6- تفسير معاملات النموذج واختبار معنوياتها.

لاختبار المعنويات يجب حساب مصفوفة تباين وتغاير التقديرات حتى يمكن حساب الأخطاء المعيارية لهذه التقديرات، ومصفوفة التباين والتغاير هي

$$\Sigma_{\hat{\boldsymbol{\beta}}} = MSE\,(\mathbf{x}'\mathbf{x})^{-1} = 0.198\begin{pmatrix} 1.208 & -0.105 & 0.072 \\ -0.105 & 0.024 & -0.047 \\ 0.072 & -0.047 & 0.124 \end{pmatrix} = \begin{pmatrix} 0.239 & -0.021 & 0.014 \\ -0.021 & 0.005 & -0.009 \\ 0.014 & -0.009 & 0.024 \end{pmatrix}$$

وبأخذ الجذر التربيعي الموجب لكل عنصر من عناصر القطر الرئيسي نحصل على الخطأ المعياري لتقدير المعامل $\mathrm{S.}E_{\hat{\beta}}$

$$\begin{pmatrix} \mathrm{S.}E_{\hat{\beta}_0} & & \\ & \mathrm{S.}E_{\hat{\beta}_1} & \\ & & \mathrm{S.}E_{\hat{\beta}_2} \end{pmatrix} = \begin{pmatrix} \sqrt{0.239} & & \\ & \sqrt{0.005} & \\ & & \sqrt{0.024} \end{pmatrix}$$

$$= \begin{pmatrix} 0.489 & & \\ & 0.068 & \\ & & 0.157 \end{pmatrix}$$

ولاختبار معنوية المعامل ، يأخذ الفرض العدم والبديل الصورة التالية:

الفرض العدم: المتغير المستقل ليس له أثر معنوي على المتغير التابع $H_0 : \beta = 0$

الفرض البديل : المتغير المستقل ذو أثر معنوي على المتغير التابع $H_1 : \beta \neq 0$

ويحسب إحصائية اختبار معنوية المعامل بالمعادلة التالية: $t^* = \hat{\beta} / \mathrm{S.}E_{\hat{\beta}}$

والجدول التالي قيم تقديرات المعاملات ، والخطأ المعياري، وإحصائية الاختبار، وفترات الثقة 95% للمعاملات .

				فترة ثقة 95%		
$\hat{\beta}$	$\mathrm{S.}E_{\hat{\beta}}$	$t^* = \hat{\beta} / \mathrm{S.}E_{\hat{\beta}}$	*Sig.*	$t_{(9, 0.975)}$	Lower	Upper
-1.088	0.489	-2.226	0.0530	2.262	-2.193	0.018
0.325	0.068	4.763	0.0010	2.262	0.170	0.479
0.387	0.157	2.472	0.0354	2.262	0.033	0.741

تفسير فترات الثقة 95% للمعاملات.

7- متوسط الإنفاق المتنبأ به لأسرة دخلها 20 ألف وعدد أفرادها 7، وما هوا حدا الثقة لهذا المتوسط، . $\alpha=0.05$

التنبؤ بقيمة المتغير التابع Y^p عند دخل 20 ألف وعدد أفراد أسرة 7 X1=20، X2=7

تقدير بنقطة ^Y

X^o=	1	20	7	
$Y^\wedge = X^o \beta^\wedge$	**8.11243**			

فترة تنبؤ للقيمة

$X^oCOV(\beta^\wedge)X^{`o}$		-0.076	0.006921	-0.00092
		0.05597		
	$S.E_{Y^\wedge}$	**0.50346**		
	$t_{(9,0.975)}$	**2.26216**		
Lower	**6.97**			
Upper	**9.25**			

التنبؤ بقيمة متوسط المتغير التابع μ^p عند دخل 20 ألف وعدد أفراد أسرة 7 X1=20، X2=7

تقدير بنقطة ^μ

$Y^\wedge = X^o \beta^\wedge$	**8.11243**

فترة تنبؤ للمتوسط

	$S.E_{\mu^\wedge}$	**0.23657**
Lower	**7.58**	
Upper	**8.65**	