

**منهجية دمج طريقي الانحدار وتحليل المركبات الأساسية في اختيار  
طاقم المتغيرات المستقلة**

**A Methodology of Emerging Regression  
And Principal Component Analyses Methods  
For Selecting Set of Variables**

د. عبد الحميد عبد المجيد البداوي  
أستاذ مساعد  
كلية المنصور الجامعية  
بغداد - العراق

: تمهد :

تتوفر عدة طرق ومنهجيات احصائية لغرض اختيار مجموعة  
المتغيرات ( التفسيرية ) التي يمكن تضمينها في النموذج الذي يتم بناءه ،  
ويأتي هذا الاختيار لاجل تحقيق هدفين حاسمين وهما :

تحاشي مشكلة العلاقات الخطية المداخلة بين المتغيرات التي يتم ترشيحها واخضاعها للتحليل وهي ما يطلق عليها بمسألة

" Multicollinearity "

والآخر هو لتقليل عدد المتغيرات التي يضمها النموذج بغية تقليل الكلفة من جهة وتسهيل عملية احتسابه واستخدامه من جهة اخرى .

ومن اهم هذه الطرق (التي سيرد التطرق اليها لاحقا) هي التي تعتمد الانحدار المتعدد Draper and ( ; Multiple Regression Principal ) ، وتحليل المركبات الاساسية ( smith 1990 Hocking 1976 Jeffers 1978, ) Component Analysis (Kendall 1980, Al- Beldawi 1978

الا ان الطريقة المقترن تطبيقها موضوع هذا البحث هي حصيلة دمج بين كلا الطريقتين اعلاه ( Daling and Tamura, 1970 ) والتي يمكن اعتبارها الحالة العملية لطريقة ( Kendall,Principal Component Regression 1959 ) . والتي من ميزاتها هو توافر مرونة عالية للباحث في اختيار المتغير الاكثر اهمية لتضمينه في النموذج ، بالإضافة الى النتائج المعنوية التي يمكن ان تتمخض عنها عملية التحليل باستخدام المنهجية المقترنة وكما افصحت عن ذلك نتائج هذا البحث .

وللوقوف على مدى أهمية المنهجية مقارنة بالطرق التقليدية الأخرى سيتم اختصار المتغيرات المرشحة لعملية التحليل باستخدام الطرق الثلاث : الانحدار المتعدد ، وتحليل المركبات الأساسية ، والطريقة المقترحة وهي الدمج بين هاتين الطريقتين ، ومن ثم توظيف معايير تقييم المعنوية لنتائج كل منها .

#### 1. مفهوم اختيار افضل طاقم من المتغيرات المستقلة

عند صياغة النموذج الخطي العام ، يفترض وجود علاقة خطية بين المتغير المعتمد  $Y$  و  $K-1$  من المتغيرات المستقلة (التفسيرية)  $(X_1, X_2, \dots, X_{K-1})$  والمتغير العشوائي  $\mu$  وان هذه العلاقة في حالة اخذ عينة من المشاهدات  $Y_i$  ( $i = 1, 2, 3, \dots, n$ ) فأن الصيغة العامة للنموذج يكون :

$$Y_i = B_0 + B_1 X_1 + B_2 X_2 + \dots + B_{K-1} X_{K-1} + \mu$$

وان هذا النموذج يجب ان يحقق بالإضافة للفرضيات المتعلقة بالمتغير العشوائي فرضية عدم وجود علاقة خطية  $\mu$  بين المتغيرات المستقلة (التفسيرية) ، اي ان هذه المتغيرات تكون مستقلة عن بعضها كما وان عدد المشاهدات التي هي  $n$  يجب ان لا تزيد على عدد المعلومات المجهولة في النموذج اي :

$$K \leq n$$

$$(33)$$

وفي حالة عدم تحقق هذه الافتراضات فإن النتائج التي يتم التوصل إليها لا تكون صحيحة ، لذلك ينبغي عدم استخدام النموذج الخطي العام إلا بعد التأكيد أولاً من مدى انطباق الافتراضات التي على أساسها تم بناء النموذج على الواقع المشاهدات . ومن بين أهم هذه الفرضيات هي المتعلقة بمشكلة العلاقات الخطية ، والتي تبرز عندما تكون هناك علاقات تامة أو شبه تامة بين المتغيرات المستقلة ، لأن قيمة محدد المصفوفة  $|x'x|$  تساوي عندئذ صفر مما يتغير معه إيجاد  $-1(x')$  وبالتالي يتغير تقدير قيم المعلمات المجهولة .

2. نموذج تحليل المركبات الأساسية Principal Component Analysis وهي طريقة احصائية وصفية تستخدم مع البيانات ذات المتغيرات المتعددة . وتقوم بتجميع كل مجموعة من البيانات المترابطة خطيا في أحد المركبات الأساسية . وتعتمد في ذلك على مصفوفة الارتباط للمتغيرات التي تعتبر المرحلة الأساس في عملية تحليل المركبات . والمركبات الأساسية  $S' Cp$  هي متغيرات عشوائية غير مترابطة ، وكل منها تضم مجموعة متغيرات عشوائية  $(X_1, X_2, \dots, X_p)$  مترابطة وتشترك باتجاه خطى ، بحيث تأخذ الصيغة التالية :

$$C_j = \sum_{i=1}^p a_{ij}, \quad j = 1, 2, \dots, p$$

(34)

وأن  $a_{ij}$  هي عبارة عن معاملات عناصر ( elements ) مصفوفة الموجهات الذاتية الطبيعية ( Normalised eigen vectors ) لمصفوفة الارتباط للمتغيرات  $X_i$ 's

وبواسطة مصفوفة قيم الموجهات الذاتية ( eigen values ) نحصل على الحجم النسبي للتغير او التباين المفسر للبيانات الاحصائية بواسطة كل من المركبات الاساسية التي يتم احتسابها من المتغيرات (orthogonal transformation)  $X_i$ 's باستخدام التحويل المتعامد (component) ويطلق على المعاملات  $a_{ij}$  عادة بتحميلات المركبة (loadings) وهذه التحميلات تشير الى وزن العلاقة بين المتغيرات  $X_i$ 's والمركبات الاساسية  $C_p$ 's بشرط اخذ الجذر التربيعي لمصفوفة ارتباط تباينات الشيوخ لقيم التباين الذاتي (Morrison, 1967) . ويطلق على قيم الجذر التربيعي لتباينات الشيوخ (Communality) ويمكن الرمز له بـ :

$$h_j = a_{j1}^2 + a_{j2}^2 + \dots + a_{jp}^2$$

### 3. نموذج الانحدار المتعدد Multiple Regression

ان جوهر الافكار التي تعتمد عليها جميع طرق الاختيار التي سيلي ذكرها هي تضمين المتغير الذي يضيف اكبر زيادة ممكنة الى قوة التفسير للنموذج ، واذا كان على المتغير ان يحذف فيجب ان يكون تأثير حذفه اقل ما يمكن على قدرة النموذج التفسيرية .

اما اهم طرق نموذج الانحدار المتعدد المستخدمة لاختيار افضل طاقم متغيرات مستقلة فهي :

- (1) طريقة شمول كافة المتغيرات (All Possible Regression) :  
وتشتمل اذا كان عدد المتغيرات ليس كبيرا ، وابرز عيوبها حاجتها لعمليات حسابية ووقت كبيرين .

- (2) طريقة الاضافة المتتالية (Forward Selection Regression) :  
وفيها اذا كانت قيمة F المجدولة هي اقل من المحسوبة عندها يتوقف البحث عن متغير ، وبعكسه يتم ادخال متغير جديد الى المعادلة واعادة الاحتساب  
 $H_0: \beta_i = 0$  أي :  
 $vs H_1: \beta_i \neq 0$

(Backward Elimination Selection) طريقة الحذف التنازلي :- وهذا اذا كانت قيمة F المحسوبة لكافة المتغيرات اكبر من قيمة F الجدولية ، عندها يحذف متغير من المعادلة والرجوع لمعرفة قيمة F المحسوبة من جديد وهكذا لغاية تفوق قيمة F الجدولية .

- (3) طريقة الخطوات المتتالية (Stepwise Selection Regression) :  
تجمع بين طريقتي الاضافة المتتالية (FS) والحذف التنازلي (BE) ، وفي كل خطوة يتم اختيار متغير ابتداء من الاكثر اهمية ولغاية عدم هبوط قيمة F المحسوبة عن قيمة F الجدولية بكلمة اخرى اجراء اختبار معاملات المتغيرات لمعرفة معنويتها من عدمها .

وتعتبر طريقة الخطوات Stepwise Regression هي اثثر الطرق استخداما وانتشارا من الناحية العملية لقلة الوقت الذي تحتاجه في عملية الاحتساب بالإضافة الى انها تعرض النتائج في كل خطوة بصورة واضحة ومرضية ومبكرة من دون الحاجة لاجراء الخطوات غير المعنية .

#### 4. منهجة الدمج بين طريقي المركبات الاساسية والاتحاد :

#### Principal Component Regression Analysis

وهي الطريقة المقترحة من قبل Kendall, 1957 ، وتمتاز بتوفيرها لمرونة كافية لاختيار المتغير الاكثر اهمية وذلك بالاعتماد على خبرة الباحث العملية ومثل هذه المرونة لا توفرها الطرق التقليدية التي تعتمد حصرا على الاساليب الميكانيكية بغض النظر عن مدى اهمية المتغير .

والفكرة التي تقوم عليها الطريقة هي : عند افتراض ان  $R$  هي مصفوفة الارتباط لـ  $P$  من المتغيرات التوضيحية او التفسيرية  $X_i$  وكما في السابق حيث ان  $(1, 2, \dots, P = I)$  وافتراضنا بأن  $C$  التي هي مصفوفة المركبات الاساسية المتعامدة  $orthogonal matrix$  و  $D$  هي المصفوفة القطرية  $Diagonal Matrix$  لقيم التباين الذاتي  $\lambda_i$  بحيث :

$$D = C^T R C$$

فأن المركبات الأساسية هي عناصر للموجة (Vector)

$$g_i = C^T X$$

وبذلك فأن انحدار المتغير التابع (الاستجابة)  $Y$  على المركبات  $g_i$  تعطي انحدار تعامدي

ويمكن تلخيص اهم الخطوات العملية للطريقة بما يلي :

- (1) اختيار عدد المركبات الأساسية التي تفسر أعلى نسبة من التباين ولنقل ما يزيد على 90% مثلا .
- (2) تحليل انحدار المتغير التابع (الاستجابة)  $Y$  مع كل مركبة يتم اختيارها في الخطوة 1 اعلاه واحتساب قيمة  $R^2$  مع التحقق من صحة اشاره المعاملات .
- (3) اختيار المركبات التي تحقق أعلى  $R^2$  .
- (4) اختيار متغير مستقل من كل مركبة تم اختيارها في الخطوة 3 اعلاه على الاسس التالية:
  - الاهمية العملية للمتغير بالنسبة للمتغير التابع (الاستجابة)
  - سهولة قياسه (كلفته)
  - درجة ارتباطه مع المتغير التابع (الاستجابة)
- (5) اجراء عملية تحليل الانحدار للمتغير المعتمد مع كل من المتغيرات المستقلة التي يتم اختيارها في الخطوة 4 اعلاه .

## 5. البيانات الاحصائية :

لأجل اختبار مستوى اداء المنهجية المقترحة ومقارنتها بمستوى اداء الطرق التقليدية فقد تم ترشيح ثلاثة عشر متغيرا توضيحا تم الحصول عليها من مسح احصائي لعينة شملت 842 مسافرا من مسافري النقل بين مراكز البلديات في العراق والتي تم توفيرها لغرض بناء نموذج احصائي لتفسير الطلب على النقل بين المدن . وسيكون المطلوب في هذا البحث هو استخدام كل من النماذج الثلاث لاختيار افضل مجموعة متغيرات معنوية من بين هذه لـ 13 متغيرا . والمتغيرات المرشحة هي :

- (1) مسافة الطريق (بالمكم)  $X_1$  قيمة مطلقة
- (2) يوم السفر  $X_2$  متغير هيكلی (1, 2, ----, 7) (Dumy Variable)
- (3) نوع واسطة النقل (باص متوسط الحجم ، باص كبير الحجم)  $X_3$  متغير هيكلی 1,2
- (4) وقت السفر(قبل الظهر ، بعد الظهر)  $X_4$  متغير هيكلی 1,2
- (5) الغرض من السفر (رحلة عمل ، رحلة غير عمل)  $X_5$  متغير هيكلی 2,1
- (6) عمر المسافر  $X_6$  قيمة مطلقة (عدد السنين)
- (7) جنس المسافر  $X_7$  متغير هيكلی 1,2
- (8) مهنة المسافر  $X_8$  متغير هيكلی 1,2,3,--5
- (9) معدل دخل اسرة المسافر الشهري  $X_9$  قيمة مطلقة (بالدينار)
- (10) طول زمن الرحلة  $X_{10}$  قيمة مطلقة (بالدفائق)
- (11) معدل وقت انتظار المسافر في المحطة  $X_{11}$  قيمة مطلقة (بالدفائق)

(12) اجور السفر X12 قيمة مطلقة (بالدينار)

(13) معدل عدد رحلات واسطة النقل الواحدة يوميا X13 قيمة مطلقة (عدد)

## 1. نتائج التحليل :

1-6 نموذج الدمج بين طريقتي المركبات الأساسية والانحدار طبقاً لمتطلبات التحليل التي تطرقنا إليها في (الفقرة 4) أعلاه ، ففي الخطوة الأولى وكما مبين في الجدول رقم (1) حصلنا على 19 مركبات أساسية معنوية استطاعت هذه المركبات أن تقوم بتفسير أكثر من 99% من التباين .  
 وعند إجراء عملية تحليل الانحدار للخطوة التالية مع كل من المركبات العشرة التي حصلنا عليها ، واحتساب قيمة  $R^2$  لكل منها ، نستدل على معنوية 7 مركبات وهي ، 10، 9، 8، 6، 5، 4، 1 وكما مبين في الجدول رقم (1) .  
 وباعتماد الاسس المشار إليها في الخطوة الرابعة لاجل تحديد متغير واحد من كل مركبة ، نجد ان مجموعة المتغيرات المختارة تشمل على التوالي كل من :  
 $X2, X5, X7, X9, X12, X11, X4$  ،  
 وعند بناء النموذج الذي سيضم هذه المجموعة من المتغيرات ، سيتحقق المعايير الاحصائية التالية :

$$R^2 = 0.84$$

$$S.E = 232$$

$$F\text{-ratio} = 609$$

$$\text{Sig at } 0.000$$

عدد المتغيرات = 7

(40)

2-6 نموذج الانحدار بطريقة كافة المتغيرات (All. Possible) وبافتراض ان البيانات خالية من مسألة العلاقات Regression) المتداخلة Multicollinearity) وتضمين المتغيرات المستقلة البالغ عددها 13 متغيرا في النموذج ، فإن النموذج سيؤول إلى ما يلي :

$$R^2 = 0.84$$

$$S.E. = 231$$

$$F = 329 \quad \text{Sig at } 0.000$$

عدد المتغيرات = 13

Table (1)

**Results of Variables Selection Using The Principal Component Regression .**

Component	R <sup>2</sup> *	Variable Selected
1	.05	X11
2	.02	---
3	.00	----
4	.05	X9
5	.06	X7
6	.04	X4
7	.04	X2
8	.00	----
9	.05	X5
10	.05	X12

\* مع حجم عينة  $n = 842$  فإن قيم  $R^2$  في الجدول اعلاه هي عالية  
المعنوية أي  $\alpha = 0.01$

3- نموذج تحليل المركبات الاساسية  
وباستخدام طريقة تحليل المركبات الاساسية ، نستدل وكما موضح في  
الجدول (2) بأن هناك 10 مركبات (عوامل) معنوية ، استطاعت تفسير 99.4%

من مجموع التباين . وبأعتماد حجم التحميل (size of loading) واهمية المتغيرات بالنسبة للمتغير المعتمد فإن هذه المركبات العشرة يمكن تسميتها كما يلي :

1. طول زمن الرحلة . وان اعلى معامل تحمل في هذه المركبة يعود لمتغيري مسافة الطريق (0.98) وطول زمن الرحلة (0.97) وان نسبة التباين التي قامت مجموعة متغيرات هذه المركبة بتفسيرها هي %33.9 .
2. واسطة النقل . وقد بلغ معامل التحميل لهذا المتغير هو 99% وساهمت المركبة الثانية بتفسير 13.7% من مجموع التباين .
3. عمر المسافر . بمعامل تحمل مقداره 97% وتفسير 8.8% من مجموع التباين .
4. معدل دخل الاسرة الشهري . بمعامل تحمل مقداره 98% وساهمت هذه المركبة بتفسير 7.9% من مجموع التباين .
5. جنس المسافر . بمعامل التحميل مقداره 99% ويتفسير 7.4% من مجموع التباين .
6. الغرض من السفر . بمعامل تحمل مقداره 99% ويتفسير 6.6% من مجموع التباين .
7. يوم السفر . بمعامل تحمل مقداره 99% ويتفسير 6.5% من مجموع التباين .
8. مهنة المسافر . بمعامل تحمل مقداره 95% ويتفسير 5.6% من مجموع التباين .

10. وقت السفر . بمعامل تحميل مقداره 95% وبتفسير 5.1% من مجموع التباين .
11. معدل وقت الانتظار في المحطة . بمعامل مقداره 93% وبتفسير 3.9% من مجموع التباين .

**Table (2)**  
**Principal Component Loading for 13 Situational Independent Variables**

Variable	Components									
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
X1	.98									
X2								.99		
X3			.99							
X4	.31									
X5						.99				.95
X6				.97						
X7						.99				
X8									.95	
X9					.98					
X10										.97
X11										.35
X12										.94
X13										.95
Eigenvalue	4.4	1.76	1.15	1.03	.97	.86	.84	.72	.66	.51
%	33.9	13.7	8.8	7.9	7.4	6.6	6.5	5.6	5.1	3.9
Variation	33.9	47.9	56.4	64.3	71.7	87.3	84.8	90.4	95.5	99.4
Cum.% Variation										

وان معايير تقييم نتائج عملية بناء النموذج الذي يتضمن مجموعة لـ 10 متغيرات المختار في اعلاه ذات اعلى تحمل Loading في كل مركبة هي :

$$\begin{array}{ll} R^2 = 0.84 \\ S.E = 231 \\ F = 0.609 & \text{Sig. at } 0.000 \end{array}$$

عدد المتغيرات = 10

لأجل اعطاء صورة واضحة من خلال المقارنة فإن الجدول التالي يعطي حصيلة كل طريقة ومستوى معنوية كل نموذج .

المعايير الاحصائية	النموذج النهائي لكافة المتغيرات بطريقة الانحدار	النموذج النهائي بالاعتماد على طريقة تحليل المركبات	النموذج النهائي بالاعتماد على طريقة الدمج بين تحليل المركبات والانحدار
عدد المتغيرات	13	10	7
R <sup>2</sup>	0.84	0.84	0.84
S.E	231	233	232
F-ratio	329	609	856
At	0.000	0.000	0.000

ومنه نلاحظ بأنه رغم ما حققته المنهجية المقترحة وهي دمج بين تحليل المركبات والانحدار في اعداد المتغيرات التي يتضمنها النموذج الا ان درجة المغفوية قد ارتفعت بالنسبة لمعيار F-ratio مع عدم حصول اي انخفاض في درجة معنوية باقي المعايير الاحصائية

## 2. الاستنتاجات والتوصيات:

### 1-7 الاستنتاجات :

1. ان النتائج التي تم خضت عن عملية التحليل وفقا للمعايير الاحصائية والاقتصادية تشير الى ان طريقة نموذج الدمج بين طريقتي تحليل المركبات وتحليل الانحدار ، تمتاز بتقليل عدد المتغيرات المستقلة الى 7 متغيرات فقط مقابل عدم الانخفاض في قدرة النموذج التفسيرية للتباين .
2. ان منهجية الدمج ساعدت في تحديد اهم المتغيرات وفقا للأسباب المنطقية التي لها علاقة في تفسير الطلب مع شفافية الكشف عن المتغيرات المهمة التي لم يتضمنها النموذج ولكنها ممثلة Proxy بواسطة المتغيرات الدالة الدالة في النموذج من خلال النظر على تحميلات المتغيرات المسطرة في مركبة عند تحليل المركبات الاساسية ، استنادا الى العلاقة الخطية التي تربط هذه المتغيرات لمركبة الواحدة .
3. ان المنهجية تزداد فائدتها ومردوداتها الاقتصادية عندما تكون امام عدد كبير من المتغيرات وذلك لما تيسره في الكشف عن تلك الاكثر اهمية من خلال التأمل بحجم معاملات التحميل لكل متغير .

4. وفقا للأسس النظرية لتحليل المركبات الأساسية القائلة بوجود علاقات خطية متراقبطة بين متغيرات كل مركبة وعدم وجود العلاقة بين المركبات . فإن عملية التخلص من العلاقات المتداخلة بين المتغيرات هي اثـرـوضوحاً وفعالية وتيسـراً من خلال اختيار الممثل المناسب في كل مركبة .

#### 7-2 التوصيات

1. بالرجوع بما تحقق من توظيف المنهجية المقترحة من نتائج مهمة ، فقد يكون من المفيد استخدام هذه المنهجية في عملية بناء النماذج الاقتصادية والاجتماعية . وعلى الأخص عند مجهولية المتغيرات التي لها تأثير فعلي على الظاهرة المدروسة بسبب عدم وجود دراسات سابقة عنها او بسبب مستجدات قد طرأت على الظاهرة مما يستدعي ترشيح اعداد كبيرة من المتغيرات لتحديد المجموعة التي لها تأثير فعلي وهـام على الظاهرة الاقتصادية والاجتماعية . فأـسـخـادـ هـذـهـ المـنـهـجـيـةـ منـ شـائـهـ تـيسـيرـ عمـلـيـةـ التـحـلـيلـ ولـلـخـرـوجـ بـشـفـافـيـةـ أـكـبـرـ فيـ اـتـخـاذـ قـرـارـ اختبار المتغيرات .

2. ان اللجوء الى المنهجية موضوعة البحث من شأنها التخلص من دوامة اختيار الطريقة المناسبة في عملية اختبار افضل طاقم متغيرات مستقلة ، والذي من شأنه محاولة اللجوء الى اغلب الطرق لاختبار الافضل وهو ما يتطلب كلفة اضافية في الوقت والاموال .

3. من المفيد ، اجراء دراسات لاحقة بأجراء تجربة محاكاة (Simulation) للوقوف على نتائج وتوصيات متقدمة ايضاً بهذا الصدد.

المصادر

1. A.H.Al-Beldawi, Methods of Scientific Research using SPSS, Dar Al-Shrouk, Aman, 2004 .
2. A.H. Al-Beldawi, Statistics for Business and Applied Sciences, Dar Al-Shrouk, Aman, 1997 .
3. Daling , J.R. and Tomura, Use of Orthogonal Factors for selection of Variables in a Regression Equation .  
An illustration , Applied Statistics, PP 260-68, 19,1970 .
4. Draper & Smith, Applied Regression Analysis, John Witley and son Inc. , London 1990 .
5. Hocking, R.R. , The Analysis and Selection of Variables in Linear Regression Biometrics, 32, PP.1-49, 1976 .
- 6.Jeffers, J.P. An Introduction to system Analysis : with ecological applications, William Clowes and sons LTD, London , 1978 .
7. Kendall , M. Multivariate Analysis, Second edition, Charts Griffin and LTD. , London, 1980.
8. Kendall , M. (1957) , A course in Mutivariat Analysis, Hafner New York , 1957 .

**9. Morrison, D.F. , Multivariate Statistical Methods,  
Mc. Graw-Hill, New York, 1967 .**

**10. Morrison, D.F., Multivariate Statistical Methods,  
Mc. Graw-Hill, New York, 1967 .**

**11. W.J. Krzanowski, Principles of Multivariate  
Analysis, Oxford University Press, 1988 .**