التحليل الإحصائي متعدد المتغيرات
(التحليل التجميعي)

"""في دراسة تحديد مستويات الهيكل الاقتصادي الاجتماعي لأسر المجتمع"""

الدكتور فريد الجاعوني
كلية الاقتصاد ـ جامعة دمشق

الدكتور عدنان غانم
كلية الاقتصاد ـ جامعة دمشق

الملخص

تتضمن هذه الدراسة تطبيق أحد أساليب التحليل الإحصائي متعدد المتغيرات هو أسلوب التحليل التجميعي الذي يعدُّ من الأساليب الإحصائية المتقدمة التي تستخدم في دراسة تحديد مستويات الهيكل الاقتصادي الاجتماعي لأسر المجتمع، ويساعد في رسم سياسات الدولة و تخطيط النشاط الاقتصادي.

لقد تم تطبيق هذا الأسلوب من خلال ستة متغيرات اقتصادية واجتماعية (متوسط الدخل، متوسط الإنفاق على التعليم، متوسط الإنفاق على الصحة ...الخ) من أجل الوصول للأهداف الآتية:

١- تحديد الهيكل الاقتصادي الاجتماعي للأسر.

٢- تحديد العوامل والمتغيرات التي تؤثر في تحديد هذا الهيكل.

٣- محاولة تطوير أساليب وتقانات التحليل الإحصائي وتطبيقها في الدراسات الاقتصادية.

مقدمة

يعتمد أسلوب التحليل الإحصائي على نواة المشكلة محل القياس والتحليل ونوع البيانات المتوافرة علما بأنَّ أسلوب التحليل الذي يستخدم لدراسة مشكلة معينة قد لا يكون مناسباً لدراسة مشكلة أخرى وذلك لاختلاف طبيعة المشكلة ونوع البيانات.

ونظراً لتعدد أهداف البحوث والدراسات وتعدد متغيراتها الاقتصادية والاجتماعية، فقد يتم تحديد الأساليب الإحصائية، التي تتفق مع طبيعة البيانات وتحقق أهداف البحوث والدراسات والتي من أهمها أسلوب التحليل الإحصائي متعدد المتغيرات "Multivariate statistical analysis " والذي يتضمن مجموعة من الأساليب أهمها:

١- أسلوب تحليل المكونات الرئيسية .Principal components Analysis.

٢- أسلوب التحليل لعاملي Factor analysis.

٣- أسلوب التحليل التجميعي Cluster Analysis .

مع العلم أن أسلوب التحليل الإحصائي متعدد المتغيرات بأساليبه المختلفة يعتمد على وصف وتحليل الظواهر ذات الأبعاد والمتغيرات المتعددة، فإذا كانت المشاهدات [x1 , x2 ..... Xp] تشترك فيما بينها بمجموعة من الخصائص والصفات بدرجات متفاوتة، فإنَّ التحليل الإحصائي متعدد المتغيرات يتناول دراسة بيانات تلك المشاهدات والتعبير عنها من خلال أكثر المتغيرات تأثيراً في الظاهرة محل الدراسة.

وسيتم من خلال هذا البحث التركيز على أسلوب التحليل التجميعي (Cluster Analysis) على اعتبار أن هناك الكثير من الدراسات والبحوث تناولت أسلوب المكونات الرئيسية، وأسلوب التحليل العاملي، في دراسة تحديد الهيكل الاقتصادي والاجتماعي لأسر المجتمع، وندرة الدراسات التي تناولت أسلوب التحليل التجميعي في تحديد مستويات الهيكل الاقتصادي والاجتماعي لأسر المجتمع.

تبدو أهمية الدراسة كمحاولة للخروج من دائرة الوصف النظري إلى بعد أشمل وأعم، يتضمن استخدام الأساليب الإحصائية المتقدمة، كأسلوب التحليل التجميعي.

أهمية البحث:

تعدُّ مشكلة تحديد المستويات الاقتصادية والاجتماعية والديمغرافية للمجتمعات من أهم المشاكل التي تواجه الحكومات على مختلف مذاهبها الفكرية وأنظمتها السياسية في وضع برامج التنمية الاقتصادية والاجتماعية الشاملة بصورة أكثر واقعية، لذلك فإنَّ دراسة وتحديد تلك المستويات ومعرفة العوامل والأبعاد التي تؤثر في تحديدها، تحتل أهمية كبيرة في عملية التخطيط للتنمية.

وتكمن أهمية تحديد المستويات، كأداة مهمة تساعد المخططين على رسم سياسات الدولة:

١- **السياسة الضريبية**: وذلك فيما يتعلق بالأجور والمرتبات، والإعانات الحكومية، ومسألة إعادة توزيع الدخل، بما يحقق العدالة الاجتماعية.

٢- **السياسة السكانية**: من حيث معدلات المواليد والوفيات، الرعاية الصحية للأم والطفل، تامين الخدمات والمستلزمات الصحية المختلفة.

٣- **السياسة التعليمية والبحث العلمي**: في وضع برامج محو الأمية والتعليم الإلزامي والتعليم المهني والجامعي، مما يساعد على وضع الخطط الكفيلة لضمان مستقبلهم المهني والوظيفي.

إن تحديد المستويات الآنفة الذكر لأسر المجتمع تعدُّ من الركائز الأساسية الهامة التي تساعد على التخطيط العلمي السليم باعتباره صورة فوتوغرافية تساعد المخططين والمنفذين على استقراء الواقع الاقتصادي والاجتماعي والديمغرافي.

ويمكن تلخيص أهداف البحث في النقاط الآتية:

١- تحديد الهيكل الاقتصادي والاجتماعي للأسر.

٢- تحديد العوامل والمتغيرات التي تؤثر في تحديد المستوى الاقتصادي والاجتماعي للأسر.

٣- محاولة تطوير أساليب وتقانات التحليل الإحصائي و تطبيقها في الدراسات الاقتصادية.

مفهوم التحليل التجميعي Inspect Cluster Analysis

يعدُّ أسلوب التحليل التجميعي "Cluster Analysis" أحد الأساليب الإحصائية المتعددة المتغيرات الذي بدأ استخدامه في السنوات الأخيرة، وأهم ما يميز هذا الأسلوب هو قدرته على تصنيف المشاهدات "Observation" في مجموعات مختلفة كل مجموعة طبقاً لخصائصها المميزة. وهناك عدة طرق تستخدم في التحليل التجميعي وتعمل هذه الطرق -رغم اختلافها - بمرحلية التجميع نفسها.

ويمن وضع بيانات المشاهدات الخاضعة للتحليل التجميعي في شكل مصفوفة على الصورة الآتية ([1]):

$$X_{(nxp)} = \begin{pmatrix} X11 & X12 & .. & X1j & ... & X1p \\ X12 & X22 & ... & X2j & ... & X2p \\ . & . & . & . & . & . \\ Xi1 & Xi2 & ... & Xij & ... & Xip \\ . & . & . & . & . & . \\ Xn1 & Xn2 & ... & Xnj & ... & Xnp \end{pmatrix}$$

حيث:

---

١ - Dubrov A. (1992) " Applied Multivariate Data Analysis ", Statistica , Moscow.

i= 1, 2, ……, n ; j = 1,2 , …… ,p

Xij= عبارة قيمة المتغير (j) بالنسبة للمفردة (i).

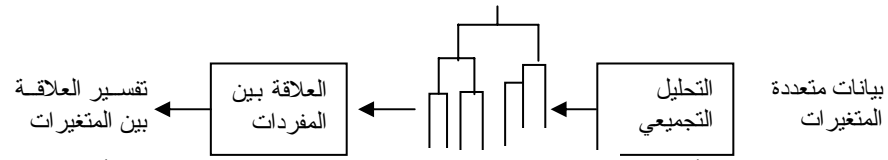
n = عدد المشاهدات الخاضعة للتحليل.

p =عدد المتغيرات الداخلة في الدراسة.

إن أسلوب التحليل التجميعي " Cluster Analysis " بطرقه المختلفة يبدأ بحساب إما مصفوفة التشابه "Similarity Matrix"، أو مصفوفة الاختلاف " Dissimilarity Matrix " بين المفردات بعضها مع بعضها الآخر، أو بين مفردة واحدة ومجموعة من المفردات.

وتبدأ عملية التحليل التجميعي بضم تلك المفردات التي تحقق أكبر قيمة تجانس فيما بينها "Within Groups". في مجموعة واحدة، وفي الوقت نفسه تحقق أقل قيمة تجانس بين هذه المجموعة والمجموعات الأخرى "Between Groups" . وفي هذا الإطار تستمر عملية التجميع وإعادة التجميع بحيث تتشكل في النهاية عدة مجموعات تشتمل كل منها على عدد من المفردات التي تتميز بترجيح أهمية خاصة معينة أو مجموعة من الخصائص.

وتتبلور نتائج عملية التحليل التجميعي " Cluster analysis" في صورة هيكل شجره يطلق عليها " dendrogram" تشتمل على كل المجموعات التي انتهت إليها عمليات التحليل. ويمكن وضع تصور للخطوات التي يمر بها أسلوب التحليل التجميعي باستخدام المخطط الآتي:



إن تجميع المفردات أو المشاه Dendrogram ما يتوقف على قياس التقارب أو التباعد بين المفردات. وهناك عدة مق قياس هذا التقارب أو التباعد يطلق عليها مقاييس التماثل أو التباعد بين ،سعرب وسوب يتم سولها على النحو الآتي:

أولاً: مقاييس التشابه Similarity measures

يتطلب التوصل إلى هيكل تجميعي بسيط من مجموعة بيانات مختلفة،إيجاد مقياس للتماثل، وذلك لقياس درجة العلاقة والتشابه بين المفردات عن طريق مقارنة كل زوج من المفردات كلاً على حده بناءً على قيم المتغيرات المشتركة، ويعتمد اختيار أي مقياس للتماثل على الحكم الشخصي للباحث بدرجة كبيرة.

وفي كثير من التطبيقات العملية لأسلوب التحليل التجميعي قد تتضمن بيانات المفردات أنماطاً مختلفة من المتغيرات الخاضعة للقياس والتحليل، فقد تكون هذه المتغيرات إما في صورة متغيرات كمية " Quantitative Data"تأخذ قيماً مختلفة، أو قد تكون في صورة متغيرات نوعية " Qualitative Data" تأخذ بدائل مختلفة أو قد تكون متغيراتٍ ثنائية " Binary Data " تأخذ القيم . (1 or 0) ولمعالجة مثل هذه البيانات يكمن تحديد قيم معاملات التماثل (S_{ijk}) لأنماط البيانات المختلفة على النحو **الآتي**:

أ- البيانات الثنائية Binary Data

لقياس درجة التماثل بين المفردات في حالة البيانات الثنائية، يتم غالباً مقارنة أزواج المفردات بناءً على وجود أو عدم وجود خصائص معينة. فالمفردات المتماثلة يكون لها خصائص مشتركة أكثر من المفردات غير المتماثلة. ومن الناحية الرياضية، يمكن وصف وجود أو عدم وجود خاصية ما باستخدام متغير ثنائي " Binay vairiable " يأخذ القيمة (1) إذا كانت الخاصية موجودة، والقيمة (0) إذا لم تكن موجودة. فمثلاً، إذا كان لدينا أربعة متغيرات ثنائية وتم قياس هذه المتغيرات بناءً على الخاصية (X_k)، فإنَّ قيم هذه المتغيرات عند المفردة (i) والمفردة (j) تكون كما يوضحه الجدول الآتي:

جدول (١) يبين تحديد قيمة (S_{ijk}) ي حالة البيانات الثنائية

قيمة المتغير (X_k)				
	(1)	(2)	(3)	(4)
المفردة (i)	1	1	0	0
المفردة (j)	1	0	1	0
قيمة (S_{ijk})	1	0	0	0

في هذه الحالة نجد أن هناك زوجين من القيم غير متماثلين هما (1 , 0) , (0 , 1) ونجد زوجين متماثلين هما (1 , 1) , (0 , 0)

وبفرض أنَّ (X_{ik}) تشير إلى خاصية المتغير الثنائي (K) بالنسبة للمفردة (i) وأن (X_{jk}) تشير إلى خاصية المتغير الثنائي (k) بالنسبة للمفردة (j)، حيث تأخذ (k) القيم P , ... 2 , 1 .

وحتى يمكن معاملة أزواج القيم المتماثلة 1,1) بطريقة مختلفة عن أزواج القيم المتماثلة (0 , 0) ، اقترحت طرق عديدة لتعريف وقياس معاملات التماثل "Similarity coefficients " بين المفردات. ولعرض هذه الطرق يمكن وضع تكرارات أزواج القيم المتماثلة وغير المتماثلة للمفردة (i) والمفردة (j) في جدول الاقتران الآتي[٢]:

جدول (٢) حالات الاقتران للمفردتين (I,j)

	المفردة j		المجموع
1	a	b	a + b
0 المفردة i	c	d	c + d

٢- Adamev B. Kildishev G. " Multiva Riate Statistical analysis Method To determine the and Economical structure " MesI , Moscow., 1995

P= a+ b+ c+ d	b+d a+c	المجموع

وتمثل a، عدد الأزواج المتماثلة (1,1)، في حين تمثل b عدد الأزواج غير المتماثلة (1,0) وهكذا. والجدول (٣) يبين بعض معاملات التماثل الشائعة والمعرفة بدلالة التكرارات الموجودة في جدول (٢).

جدول(٣) معاملات التشابه لجميع المفردات

التفسير	المعامل
أوزان متساوية للأزواج المتشابهة (1,1)،(0 ,0)	(1) $Sij = \frac{a+d}{p}$
وزن مضاعف للأزواج المتشابهة (1,1) و (0 , 0)	(2) $Sij = \frac{2(a+d)}{2(a+d)+b+c}$
وزن مضاعف للأزواج غير المتماثلة	(3) $Sij = \frac{a+d}{a+d+2(b+c)}$
استبعاد الأزواج (0,0) من البسط	(4) $Sij = \frac{a}{p}$
استعباد الأزواج (0,0) من كل من البسط والمقام	(5) $Sij = \frac{a}{a+b+c}$
استبعاد الأزواج (0,0) من البسط والمقام وإعطاء وزناً مضاعفاً للأزواج (1,1)	(6) $Sij = \frac{٢a}{2a+b+c}$
استبعاد الأزواج (0,0) من البسط والمقام وإعطاء وزناً مضاعفاً للأزواج غير المتماثلة.	(7) $Sij = \frac{a}{a+2(b+c)}$
نسبة الأزواج المتماثلة إلى غير المتماثلة بعد استبعاد الأزواج (0,0)	(8) $Sij = \frac{a}{b+c}$

ولتوضيح ذلك نأخذ عينة فرضية مكونة من ٤ أسر، تتضمن عدة متغيرات اقتصادية واجتماعية:

جدول (٤) :حالات الاقتران للمتحولات المدروسة

الأسر	الحالة العملية (X1)	حالة المسكن (X2)	متوسط الدخل الشهري (X3)	متوسط الإنفاق على التعليم (X4)	متوسط الإنفاق على الخدمات الصحية (X5)	امتلاك سيارة (X6)
١	يعمل	ملك	٩٠٠٠	١٠٠٠	٨٠٠	يمتلك
٢	يعمل	إيجار	١١٠٠٠	١٨٠٠	١١٠٠	لا يمتلك
٣	لا يعمل	إيجار	١٠٠٠٠	١٧٠٠	١٦٠٠	يمتلك
٤	يعمل	إيجار	٥٠٠٠	١٨٠٠	٥٠٠	يمتلك

حيث يمكن تعريف المتغيرات الثنائية [X1, x2, x3, x4, x5, x6] كما يلي:

الحالة العملية

$$X1=\begin{cases}1 & \text{يعمل}\\ 0 & \text{لا يعمل}\end{cases}$$

متوسط الإنفاق على التعليم

$$X4=\begin{cases}1 & \geq ١٧٠٠\\ 0 & < 1700\end{cases}$$

حالة المسكن

$$X2=\begin{cases}1 & \text{ملك}\\ 0 & \text{إيجار}\end{cases}$$

متوسط الإنفاق على الصحية

$$X5=\begin{cases}1 & \geq ١٠٥٠\\ 0 & < 1050\end{cases}$$

امتلاك سيارة:

$$X6=\begin{cases}1 & \text{يمتلك}\\ 0 & \text{لا يتملك}\end{cases}$$

متوسط الدخل

$$X3=\begin{cases}1 & \geq ١٠,٠٠٠\\ 0 & < 10,000\end{cases}$$

ويتم وضع قيم تلك المتغيرات التي تم الحصول عليها للمفردات المختلفة في جدول في الصورة الآتية:

جدول (٥) قيم المتغيرات الثنائية التي تم الحصول عليها

الأسر	X_1	X_2	X_3	X_4	X_5	X_6
١	١	١	٠	٠	0	١
٢	١	٠	١	١	١	٠
٣	٠	٠	١	١	١	١
٤	١	٠	٠	٠	٠	١

والجدول (٦) يبين عدد الأزواج المتماثلة وغير المتماثلة للمفردتين (2 , 1)

جدول (٦) عدد الأزواج المتشابهة وغير المتشابهة للمفردتين (1,2)

المجموع	المفردة z	
	١	٠
١	١	٢

٣	٠ ٣	المفردة ١ ٠
٦	٢ ٤	المجموع

وباستخدام معامل التشابه (1)، الذي يعطى أوزاناً متساوية للأزواج المتماثلة وغير المتماثلة نجد أن:

$$S_{ij} = \frac{a+b}{P} \qquad 0 \le S_{ij} \le 1 \qquad (1)$$

ويأخذ معامل التشابه القيم بين (1 , 0)، ومن ثَمَّ فإنَّ معامل التماثل للمفردتين الأولى والثانية يكون في الصورة الآتية:

$$S_{ij} = \frac{1+0}{6} = \frac{1}{6}$$

وكذلك يتم بعد ذلك حساب قيم معامل التشابه (١)، لجميع المفردات الأخرى، وتوضع هذه القيم في مصفوفة متماثلة أبعادها ٤ ×٤ كما في الصورة الآتية:

الأسر

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & & & \\ \frac{1}{6} & 1 & & \\ \frac{3}{6} & \frac{3}{6} & 1 & \\ \frac{2}{6} & \frac{3}{6} & \frac{4}{6} & 1 \end{bmatrix}$$

وبناءاً على قيم معامل التماثل المستخدم نستنتج أنَّ الأسرتين الثالثة والرابعة هما أكثر الأسر تماثلا وأن الأسرتين الأولى والثانية اقلهما تماثلاً في حين تقع أزواج الأسر الأخرى بينهما. فإذا تم تقسيم الأسرة إلى مجموعتين متماثلتين بناءً على قيم معامل التماثل المستخدم، فإنَّه يمكن تكوين المجموعة (٤،٣) والمجموعة (١،٢).

ب- البيانات النوعية Qualitative Data:

في حالة البيانات النوعية التي تأخذ بدائل مختلفة، فإنَّ قيمة معامل التشابه (S_{ijk}= 1) إذا كانت الأسرة (j) والأسرة (i) لهما الخاصية نفسها (X_k)، ويأخذ القيمة (0) في حالة اختلاف الأسرتين. فمثلاً، إذا كان لدينا أربعة متغيرات وصفية هي، المستوى الاقتصادي والاجتماعي، الحالة الاجتماعية، لون الشعر ولون العيون، وكانت بياناتها كما يأتي:

١- المستوى الاقتصادي والاجتماعي ((١) مستوى منخفض، (٢) مستوى متوسط، (٣) مستوى مرتفع)

٢- الحالة الاجتماعية ((١) أعزب، (٢) متزوج، (٣) مطلق).
٣- لون الشعر (١) أسود، (٢) أصفر، (٣) بنى)
٤- لون العيون ((١) سوداء، (٢) خضراء، (٣) زرقاء).
فإنَّ قيم تلك المتغيرات للأسرة (i) والأسرة (j) تكون كما يوضحه الجدول الآتي: [٣]

جدول (٧) يبين تحديد قيمة (S_{ijk})، للبيانات النوعية

	قيمة المتغير (X_k)			
	(1)	(2)	(3)	(4)
الأسرة (i)	١	١	٢	٢
الأسرة (j)	1	2	1	2
قيمة (S_{ijk})	١	٠	0	1

جـ- البيئات المختلطة Mixture:

في كثير من التطبيقات العملية لأسلوب التحليل التجميعي قد تتضمن بيانات المفردات أو الأسر أنواعاً مختلفة من البيانات، الخاضعة للقياس والتحليل، ولمعالجة مثل هذه البيانات قدم جور "gower1971" صيغة رياضية، يطلق عليها معامل جور للتماثل " Gower's Similarity Coefficient" حيث يقيس هذا المعامل درجة التماثل بين المفردات بناء على خليط من الأنواع المختلفة لبيانات المتغيرات الخاضعة للقياس، حيث يأخذ معامل جور للتماثل العلاقة الرياضية الآتية [٤]

$$Sij = \sum_{k=1}^{P} S_{ujk} \Big/ \sum_{k=1}^{P} W_{ijk}\,,\ 0 \le Sij \le 1,\ (i,j) = 2,....,n \qquad (2)$$

حيث:

٣- Romesburg, H. Charles, (1984) " Cluster Analysis for Researchers " , Lifetime Learning Publications, NewYork , P. 142.
٤- Mhitarian V. " Multivariate statistical Analysis " MesI, Moscow - 1988

S_{ij} = معامل التشابه بين المفردة (i) والمفردة (j) ، وتأخذ القيم بين (1 , 0)، حيث تأخذ القيمة (١) في حالة التشابه التام بين المفردتين محل الدراسة، وتأخذ القيمة (٠) في حالة الاختلاف التام بين المفردتين، كما أنها تأخذ كسراً موجباً فيما بين تلك القيمتين.

S_{ijk} = معامل التشابه بين المفردة (i) والمفردة (j) بالنسبة للمعيار (X_k)

W_{ijk}= أوزان المتغيرات.

P = عدد المتغيرات أو المعايير الداخلة في قياس درجة التشابه.

وقد افترض " Gower " أن قيمة معامل التشابه ($S_{ijk} = 1$) عندما تكون هناك درجة معينة من التماثل بين المفردتين بناءً على المعيار (X_k)، ويأخذ القيمة (0) عندما يكون هناك اختلاف أو عدم إمكانية المقارنة بين المفردتين محل الدراسة.

أما بالنسبة لقيم أوزان المتغيرات (W_{ijk})، فإنَّها تأخذ القيمة (1) إذا كانت هناك إمكانية للمقارنة بين المفردة (i) والمفردة (j) بناءً على المعيار (X_k , k =1 ,2,… p)، كما أنها تأخذ القيمة (٠) عندما لا تكون هناك إمكانية المقارنة بين المفردتين محل الدراسة.

د- البيانات الكمية Quantitative Data

تم الإشارة فيما سبق بأنه يتم معالجة وحساب معاملات التماثل وفقاً لطبيعة البيانات الخاضعة للقياس والتحليل، وعندما يتضمن التحليل بيانات كمية فإنَّه يتم حساب مقياس التماثل (S_{ijk}) بين المفردتين (i , j) في الصورة الآتية [٥]

$$S_{ijk} = 1 - \left| X_{ik} \;\; X_{jk} \right| / Rk \quad X_{ik} = X_{ik} / R_k \qquad (٣)$$

حيث:

S_{ijk} = قيمة معامل التماثل بين المفردتين (i, j) بالنسبة للمتغير (X_k).

X_{ik} = قيمة المتغير (X_k) بالنسبة للمفردة (i)

X_{jk} = قيمة المتغير (X_k) بالنسبة للمفردة (j)

R_k = مدى المتغير (X_k) بالنسبة للمفردة (I) والمفردة (j)

فإذا كانت قيم كل من (X_{ij} , X_{jk}) متساويتين، فإنَّ قيمة معامل التماثل .(S_{ijk} = I) أم إذا كانت قيم كل من (X_{ik}, X_{jk}) غير متساويتين، فإنَّ قيمة معامل التماثل (S_{ijk}) تصل إلى أدنى قيمة لها، حيث تكون قيمتها في هذه الحالة كسراً موجباً.

٥ - Krzanowski, W.J. (1995) , OP , Cit.

ثانياً: مقاييس الاختلاف Dissimilarity measures

يعدُّ معامل الاختلاف "Euclidean Distance Coefficient , d_{ij}" من أهم وأكثر المقاييس استخداماً في الحياة العملية لدراسة الخواص المشتركة بين المفردات ومدى اقترابها أو تباعدها من بعضها بعضاً، حيث يعرف معامل الاختلاف (d_{ij}) بين المفردة (i) والمفردة (j) في الصورة الآتية:[٦]

$$(4)\quad Sij = \left\{ \sum_{K=1}^{P} (X_{ik} - X_{jk})^2 \right\}^{\frac{1}{2}} \qquad 0 \le dji < \infty$$

حيث:

Xik = القيمة التي يأخذها المتغير (K) بالنسبة للمفردة (i).

Xjk = القيمة التي يأخذها المتغير (K) بالنسبة للمفردة (j).

dij = معامل التباعد بين المفردتين (j ، i).

ويتم استخدام العلاقة (4) لاشتقاق مقياس التشابه بين المفردات المختلفة – فمثلاً، إذا كان التباعد بين المفردة (i) والمفردة (j) هو (d_{ij}) فإنَّ مقياس التشابه (Sij) بين هاتين المفردتين يكون في الصورة الآتية ([7]).

$$\tilde{S}_{ij} = \frac{1}{1 + d_{ij}} \qquad 0 \le Sij \le 1 \qquad (5)$$

وفي المقابل وطبقاً للاقتراح الذي قدمه "Gower, 1966 يمكن أيضاً اشتقاق مقاييس الاختلاف من مقاييس التشابه إذا كانت مصفوفة التباعد غير أكيدة السالبية " Nonnegative definite" فإذا كانت مصفوفة المسافة غير أكيدة السالبية وتمت قسمة قيم مقياس التشابه المستخدم على مقدار ثابت لا يساوي الصفر بحيث أن Sij = 1، فإنَّه يمكن اشتقاق معامل الاختلاف من معامل التشابه

في الصورة الآتية:

$$\tilde{d}_{ij} = \sqrt{2(1 - \tilde{S}_{ij})} \qquad (6)$$

اختبار الفروق المعنوية لمتوسطات المجموعات

٦- Krzanowski , W.J. (1995)., OP.Cit.

٧ - المرجع نفسه.

بعد أن تم تحديد المستويات الاقتصادية والاجتماعية للهيكل البنائي الاقتصادي والاجتماعي لأسر المجتمع، فإنَّه يتطلب اختباراً أكانت هناك فروق معنوية بين متوسطات المجموعات التي تم التوصل إليها أم لا، وفي هذا الجانب هناك عدة اختبارات من أهمها ما يطلق عليه اختبار Wilk's Lambda Test, Λ"[٨]، حيث Λ

حيث إنَّ:

det= " Determinant matrx" محددة المصفوفة

W= " Within – groups' matrix" مصفوفة الفروقات داخل كل مجموعة

T= W+B

B= " between – groups' matrix " مصفوفة الفروقات ما بين المجموعات

ويكون الاختبار الإحصائي لمعنوية الفروق في الصورة الآتية:

$$\Lambda = \frac{\det(W)}{\det(T)} \qquad (7)$$

$H_0= \mu_1 = \mu_2 \ldots = \mu I$: *فرضية العدم*

$H_1 = \mu_1 \neq \mu_2 \ldots \neq \mu I$ *الفرضية البديلة:*

Where μ_1 *is the multivariate mean of the ith group.*

حيث يتحول "Wilk's Lambda Λ إلى " F- Ratio" كما في الصورة الآتية:

$$F \frac{1-\Lambda^{\frac{1}{2}}}{\Lambda^{\frac{1}{2}}} \frac{N-P-2}{P} \qquad (8)$$

حيث:

N -العدد الكلي للمفردات مع درجتي حرية. V=2P

$$\nu_2 = 2\,(N - p - 2)$$

حيث p هو عدد المتغيرات المدروسة.

فإذا كانت نتيجة الاختبار هو رفض (H_0)، يمكن الاستنتاج أن هناك فروقاً معنوية بين متوسطات المستويات للظاهرة المدروسة.

الخاتمـــة:

٨- Cooley , W.W , (1971) "Multivariate Data Analysus ' , Johi Wiley & Sons , Ins. , New Yourk,.P437.

بعد أن يتم تحديد المستويات للظاهرة المدروسة فإنَّ الخطوة الآتية تتطلب من الباحث توصيف وتوزيع مفردات الدراسة داخل هذه المستويات ويتم باستخدام أسلوب التحليل التمييزي Discriminate Analysis الذي يعدُّ من أحد أساليب التحليل الإحصائي متعدد المتغيرات ويهدف إلى بناء قاعدة، تمكن الباحث من الفصل بين المستويات المحددة بواسطة أسلوب التحليل التجميعي، والتمييز بين مجموعتين أو أكثر من المشاهدات التي يتم جمعها من مجتمعات مختلفة بناء على خصائص تلك المفردات.

وبشكل عام يمكن إيجاز الجوانب الأساسية للتحليل التمييزي في الآتي:

١- دراسة مدى التداخل بين المجتمعات.

٢- تكوين قاعدة للفصل بين المجتمعات محل البحث والدراسة بناء على عينة من المفردات يتم سحبها من هذه المجتمعات.

٣- على الأغلب يكون مجتمعان معلومين مقدماً أو أكثر تتشابه في خصائصها ولكن منفصلة كمياً.

ويمكن دراسة هذا الأسلوب من التحليل الإحصائي متعدد المتغيرات بشكل أوسع وأشمل في دراسة منفصلة.

المراجـــــع

1. Adamev B. Kildishev G. " Multiva Riate Statistical analysis Method To determine the and Economical structure " MesI , Moscow., 1995

2. Cooley , W.W , (1971) “Multivariate Data Analysus ‘ , Johi Wiley & Sons , Ins. , New Yourk,.P437.

3. Dubrov A. (1992) “ Applied Multivariate Data Analysis “, Statistica , Moscow.

4. Krzanowski, W.J. (1995) Moscow.

5. Makhtirian V. " Multivariate statistical Analysis " Mecu, Moscow - 1988

6. Romesburg, H. Charles, (1984) " Cluster Analysis for Researchers " , Lifetime Learning Publications, NewYork , P. 142.

تاريخ ورود البحث إلى مجلة جامعة دمشق ٢٠٠٠/٣/٢٢.