التحليل الإحصائي متعدد المتغيرات (التحليل التمييزي) (في توصيف وتوزيع الأسر داخل الهيكل الاقتصادي الاجتماعي في المجتمع)

الدكتور فريد الجاعوني الدكتور عدنان غانم

كلية الاقتصاد

جامعة دمشق

الملخص

يتضمن هذا البحث دراسة أحد أساليب التحليل الإحصائي متعدد المتغيرات هو أسلوب التحليل التمييزي الذي يعد من الأساليب الإحصائية المتقدمة التي تستخدم في توصيف وتوزيع الأسر داخل الهيكل الاقتصادي الاجتماعي للمجتمع، ويساعد في رسم خطط التنمية الاقتصادية والاجتماعية التي تهدف إليها الدولة والوقوف على أنسب الطرائق من حيث عدالة توزيع الدخل والعبء الضريبي والإعانات الحكومية لأسر المجتمع بصورة أكثر واقعية.

يطبق هذا الأسلوب من خلال عدة متغيرات تُحدد من قبل الباحث من أجل الوصول إلى الأهداف الآتية:

- توصيف وتوزيع الأسر داخل الهيكل الاقتصادي للمجتمع.
- تحديد العوامل والمتغيرات التي تؤثر في توصيف وتوزيع الأسر داخل الهيكل الاقتصادي الاجتماعي في المجتمع.
- محاولة تطوير أساليب وتقانات التحليل الإحصائي وتطبيقها في الدراسات الاقتصادية.

I – مقدمة:

يعتمد أسلوب التحليل الإحصائي على نوع المشكلة محل القياس والتحليل، ونوع البيانات سواء كانت بيانات (نوعية، أم كمية) لذلك فالتحليل الذي يستخدم لدراسة مشكلة معينة قد لايكون مناسباً لدراسة مشكلة أخرى.

يعتمد أسلوب التحليل الإحصائي متعدد المتغيرات بأساليبه المختلفة على وصف وتحليل الظواهر ذات الأبعاد والمتغيرات المتعددة. فإذا كانت المشاهدات $[x_1, x_2, \ldots x_p]$ **تشترك فيما بينها بمجموعة من الخصائص والصفات بدرجات متفاوتة، فإن التحليل الإحصائي متعدد المتغيرات يتناول دراسة بيانات تلك المشاهدات والتعبير عنها من خلال أكثر المتغيرات تأثيراً في الظاهرة محل الدراسة.**

سيتم من خلال هذا البحث التركيز على أسلوب التحليل التمييزي (Discriminant Analysis) **على اعتبار أن هناك دراسات سابقة تناولت الأساليب الإحصائية الأخرى في تحديد الهيكل الاقتصادي والاجتماعي لأسر المجتمع، مع ندرة الدراسات التي تناولت أسلوب التحليل التمييزي في توصيف وتوزيع الأسر داخل الهيكل الاقتصادي الاجتماعي في المجتمع.**

II – مشكلة البحث:

تتمثل مشكلة البحث في هذه الدراسة في تحديد الهيكل الاقتصادي والاجتماعي للأسر، وتحديد المتغيرات الفاعلة والمؤثرة في تحديد ذلك الهيكل، وذلك من خلال تحليل وقياس العديد من الأبعاد الاقتصادية والاجتماعية.

وبالتحديد فإن هذه الدراسة تقوم على أساس وجود تساؤلات حول: ما المستويات الاقتصادية والاجتماعية الموجودة في مجتمع ما؟ وما المتغيرات الاقتصادية والاجتماعية التي يمكن التركيز عليها في التمييز وتقسيم أسر مجتمع ما إلى مستويات اقتصادية واجتماعية مختلفة؟

III – أهداف البحث وأهميته:

تكمن أهمية البحث في استخدام أسلوب التحليل التمييزي كأداة مهمة تساعد المخططين على رسم سياسات الدولة (الضريبية، الدخلية فيما يتعلق بالأجور، الإعانات، توزيع الدخل في تحقيق مبدأ العدالة الاجتماعية لأسر المجتمع). والتعرف على الظروف الاقتصادية والاجتماعية التي تحيط بالأسرة في مرحلة من مراحل التغير التي يمر بها المجتمع، وتتلخص أهداف البحث فيما يأتي :

- تحديد العوامل والمتغيرات التي تؤثر في توزيع وتوصيف الأسر داخل الهيكل الاقتصادي الاجتماعي في المجتمع وتحديد المستوى الاقتصادي الاجتماعي الذي تنتمي إليه هذه الأسرة.
- محاولة إلقاء الضوء على أساليب وتقنيات التحليل متعدد المتغيرات، كي تصبح ملائمة للبيانات والمعلومات المتوافرة عن ظاهرة البحث.

V ـ فروض البحث:

لما كان الهدف من هذه الدراسة يتمثل في محاولة دراسة وتحديد المستوى الاقتصادي والاجتماعي للأسرة في المجتمع، وتحديد المتغيرات الاقتصادية والاجتماعية الفاعلة والمؤثرة فيه. ومع وضع الخصوصية الاقتصادية والاجتماعية والثقافية للمجتمع المدروس في الحسبان، أمكن صياغة الفرض الآتي:

- لا توجد فروق ذات دلالة إحصائية بين أسر المجتمع.

وللتمكن من اختبار هذا الفرض وتسهيلاً لعملية القياس، يمكن صياغته في شكل فروض فرعية قابلة للاختبار الإحصائي:

1 ـ لا توجد علاقة ارتباطية ذات دلالة إحصائية بين المستوى التعليمي والمستوى الاقتصادي والاجتماعي لأسر المجتمع.

2 ـ لا توجد علاقة ارتباطية ذات دلالة إحصائية بين المهنة والمستوى الاقتصادي والاجتماعي لأسر المجتمع.

3 ـ لا توجد علاقة ارتباطية ذات دلالة إحصائية بين الدخل، والمستوى التعليمي لأسر المجتمع.

V ـ منهجية الدراسة:

اعتمد الباحث الأسلوب الوصفي التحليلي لأسلوب التحليل التمييزي وذلك من خلال استنباط كيفية استخدام هذا الأسلوب في التحليل الإحصائي في توصيف وتوزيع الأسر داخل الهيكل الاقتصادي الاجتماعي في المجتمع، بعد الاطلاع على المراجع العلمية المتخصصة والبحوث السابقة ذات العلاقة بموضوع الدراسة، وللتعرف أيضاً على المعايير والأبعاد الاقتصادية والاجتماعية التي تحدد المستوى الاقتصادي الاجتماعي.

VI ـ مفهوم التحليل التمييزيInspect Discriminant Analysis:

بعد تحديد الهيكل الاقتصادي والاجتماعي لأسر المجتمع نبدأ بوضع قاعدة لإعادة توزيع وتوصيف الأسر داخل ذلك الهيكل الذي تنتمي إليه أسر جديدة تتم دراستها مستقبلاً، ويعد أسلوب التحليل التمييزي أحد الأساليب الإحصائية متعددة المتغيرات التي يمكن الاعتماد عليها في بناء هذه القاعدة، وذلك من خلال بيانات عينة من الأسر يتم سحبها من تلك المستويات لدراسة خصائصها الاقتصادية

والاجتماعية وتوزيعها على المستويات المختلفة المنتمية إليها، وغالباً ما يتم تحديد هذه المستويات الاقتصادية الاجتماعية مسبقاً (مرتفع، متوسط، منخفض)، ويمكن تلخيص خطوات أسلوب التحليل التمييزي فيما يأتي:

1 ـ دراسة مدى التداخل بين المجتمعات (مجتمعين أو أكثر):

يعتمد أسلوب التحليل التمييزي للفصل بين المجتمعات على أن هناك معلومات عن مجتمعين أو أكثر تتشابه في خصائصها ولكنها منفصلة كمياً.

ويفرض أن هناك (m) **من العينات العشوائية من مجتمعات مختلفة أحجامها** ($n_1,n_2,n_3…n_m$) **وأن قيم المتغيرات** ($x_1,x_2, ….. x_p$) **للعينات معلومة. فإنه يمكن وضع بيانات تلك الظاهرة القابلة للتصنيف في صورة جدول يأخذ الشكل الآتي:** (Bryanf.J, Manly 1994) .

جدول يبين شكل بيانات المفردات القابلة للتصنيف باستخدام دالة التحليل التمييزي

Individual	X_1	X_2	…	X_p	
1	X_{111}	X_{112}	…	X_{11p}	
2	X_{211}	X_{212}	…	X_{21p}	I
.	.	.	.	.	
.	.	.	.	.	
.	.	.	.	.	
n_1	X_{n111}	X_{n112}	…	X_{n11p}	
1	X_{121}	X_{122}	…	X_{12p}	
2	X_{221}	X_{222}	…	X_{22p}	
.	.	.	.	.	II
.	.	.	.	.	
.	.	.	.	.	
n_2	X_{n221}	X_{n222}	…	X_{n22p}	
.	.	.	.	.	
.	.	.	.	.	
.	.	.	.	.	
1	X_{1m1}	X_{1m2}	…	X_{1mp}	
2	X_{2m1}	X_{2m2}	…	X_{2mp}	m
.	.	.	.	.	
.	.	.	.	.	
.	.	.	.	.	
n_m	X_{nmm1}	X_{nmm2}	…	X_{nmmp}	

وإن الهدف من دراسة مدى تداخل المجتمعات هو تكوين قاعدة للفصل بين المجتمعات محل البحث بناء على عينات من المفردات يتم سحبها من هذه المجتمعات .

2 ـ التقسيم الهيكلي لأسر المجتمع باستخدام دوال التمييز:

في هذا الجانب يتم استخدام دوال التمييز لتوصيف وتوزيع مفردات الدراسة داخل هذه المستويات وتحديد المستوى الاقتصادي والاجتماعي الذي تنتمي إليه مفردة جديدة تتم دراستها مستقبلاً، حيث يتم استخدام دوال التمييز في تحديد متوسطات المستويات الاقتصادية والاجتماعية لنصل إلى تحديد المستوى الذي تنتمي إليه المفردة محل الدراسة.

ونشير هنا إلى أن العديد من الدراسات لجأ إلى استخدام معايير وأبعاد اقتصادية واجتماعية مختلفة لتحديد الهيكل الاقتصادي والاجتماعي لأسر المجتمع. وتكاد هذه الدراسات تتفق فيما بينها على اعتبار بعض معاملات المتغيرات الاقتصادية والاجتماعية المميزة للمستويات على أنها أكثر أهمية في تحديد هذه المستويات، ومن هذه المتغيرات : الدخل، المهنة، المستوى التعليمي لكل من الزوج والزوجة، خصائص المسكن، نسبة الإعالة، الممتلكات، الإنفاق على الخدمات الطبية، الإنفاق على التعليم والثقافة، الإنفاق على الغذاء والمستلزمات الشخصية، درجة الوعي الثقافي والسياسي، وكل بعد من هذه الأبعاد يسهم بقدر متباين في تحديد دالة التمييز الأكثر أهمية، وتأتي بقية المتغيرات المميزة الأخرى بدرجة متدرجة في الأهمية حسب معاملات الدالة -ولتوزيع أسر المجتمع المدروس في المستويات الاقتصادية والاجتماعية سواء كانت هذه الأسر خارج نطاق هذا الهيكل أو الأسر التي تنتقل في مستوى اقتصادي واجتماعي إلى آخر داخل هذا الهيكل، فإنه يتم استخدام قاعدة التخصيص الآتية: (Kandil, A.M.1992)

$$\sum_{j=1}^{2}\left(\hat{y}_j - \bar{y}_{kj}\right) \ , \ \ (j = 1,2);(k = 1,2,3)$$

حيث:

$\hat{y}_j$ **: الدرجات التي تأخذها المفردة الجديدة** (i)

$\bar{y}_{kj}$ **: متوسط الدرجات المميزة للمستوى** (k) **الناتجة عن دالة التمييز** (j)

k **= عدد المستويات الاقتصادية والاجتماعية الناتجة من التحليل التمييزي.**

j **= عدد دوال التمييز الناتجة من التحليل التمييزي.**

وتُستَخدم قاعدة التخصيص السابقة لتوصيف وتوزيع أسر المجتمع داخل المستويات الاقتصادية والاجتماعية التي تُشكل فيما بينها الهيكل البنائي للنسيج الاقتصادي والاجتماعي لأسر المجتمع، وذلك على النحو الآتي:

1 ـ **يتم إيجاد متوسطات الدرجات المميزة** ($\bar{y}_k$ ،k=1 ،2 ،3) **للمستويات الاقتصادية والاجتماعية.**

2 ـ **يتم إيجاد قيم** ($\bar{y}_{kj}$) **بالتعويض بمتوسطات المتغيرات المميزة لكل مستوى من المستويات الاقتصادية والاجتماعية الثلاثة في دالتي التمييز.**

3 ـ **توصيف وتوزيع أسر المجتمع المدرروس داخل المستويات الاقتصادية الاجتماعية من خلال استخدام دوال التمييز** (Discriminate Function)

في الحياة العملية وفي كثير من التطبيقات العملية لأسلوب التحليل التمييزي، قد تتضمن البيانات خليطاً من أنماط المتغيرات المستمرة، والمنفصلة والاسمية (Mixtures of continous, discrete and nominal variables)**. ومن ثم فإنه يتطلب وضع قاعدة تمكن من الفصل بين المجتمعات بناء على بيانات المفردات التي تتضمن هذه الأنماط من المتغيرات؛ وذلك وفق النموذج المقترح لتوزيع المفردات على المجتمعات المختلفة** (Kandil, A,M,1992) **:**

يفترض النموذج عدة افتراضات أهمها :

1 ـ **هناك** { m } **من المجموعات هي** $\{\prod_1, \prod_2 ....., \prod_m\}$**، يتطلب الفصل بينها بناء على بيانات تتضمن خليطاً من أنماط المتغيرات** (Mixtures of continous, discrete and nominal variables) **لـ** {n_i} **من المشاهدات، معلوم أنّها من المجموعة** $\{\prod_i , i=1,2,...m\}$

2 ـ **كل مشاهدة من مشاهدات البحث تتضمن ثلاثة أنواع من البيانات تمثلها ثلاثة متجهات مختلفة على النحو الآتي:**

أ ـ **المتجه الأول** $Z = (Z_1, Z_2, \dots , Z_g)$ **يتكون من** "g-component vector of nominal variables"**، حيث إن المتغير الاسمي رقم** r th **له** (k_i) **حالة.**

ب ـ **المتجه الثاني** $X=(x_1, x_2, \dots , x_t)$ **يتكون من** "t-component vector of discrete variables"**، يتبع التوزيع كثير الحدود بالمعالم** (n_{ixj}, ϕ_{ixj}) **، حيث:**

($i=1, 2, \dots , m; j=1,2, \dots, t;$ and $x_j = 0, 1,2, \dots,n`_j$)

ج ـ **المتجه الثالث** $Y= (Y_1, Y_2, \dots , Y_s)$ **يتكون من** S-component vector of continuous variables **يتبع التوزيع الطبيعي متعدد المتغيرات بالمعالم** (μ_{ixj}, Σ_{ixj})

وطبقاً للافتراضات السابقة، تم استبدال كـل "nominal variable" **بــ** (k_i-1) **حالــة مــن المتغيــرات الصورية** "dummy binary variables"**، حيث تأخذ هذه المتغيرات القيمة** (0)**، ماعدا المتغيرات رقم** r th (r=1, 2, … , k_i -1)**، ويلاحظ أن كل المتغيرات الصورية** "dummy binary variables" **تأخـذ القيمــة** (0) **للمتغير الاسمي في الحالة** K_i th**.**

فإذا كان هناك "g-nominal varibale" **، وكل متغير من هذه المتغيرات له** (k_i) **حالة. فإنه يتم استبدالها بـ** g(k-1) **من المتغيرات الصورية** "dummy binary variables"**، كل متغير منها يأخذ القيمة** (0 or 1)**.**

ومن ثم يتم معاملة هذه المتغيرات على أساس أنّها كثيـر الحـدود "multinomial variables" **، حيـث يمكن تكوين** $\{\ell = 2^{g(k_i-1)}\}$ **حالة من أنماط الاستجابات** (Zg=0or 1) **ضمن جدول يتكون من "** $\ell - cells$ **" لكل حالة من هذه الحالات.**

فعلى سبيل المثال، فإذا كان لدينا "Tow nominal variables" **وكل متغير من هذه المتغيرات لـه ثـلاث حالات، فإنه يتم استبدال كل** "nominal variable" **بـ** "Tow dummy binary variables"**. ومن ثم يكـون لدينا** "4-binary variables"**،** **تمثـل** (Z_1, Z_2) "first nominal variable"**،** **تمثـل** (Z_3, Z_4) "second nominal variable"**، حيث يكون لدينا** $\{\ell = 2^{g(k_1-1)} = 2^4\}$ **خليـة. والجـدول** (1) **يبـين الحـالات المختلفة التي تم الحصول عليها لكل مجموعة**:

جدول (1) : نتائج أربعة متغيرات ثنائية

Cel No.	The first nominal variable				The second nominal variable			
	Z_1		Z_2		Z_3		Z_4	
	0	1	0	1	0	1	0	1
(1)	0	0	0	0	0	0	0	1
(2)	0	0	0	0	0	1	0	0
(3)	0	0	0	0	0	0	0	0
(4)	0	0	0	0	0	1	0	1
(5)	0	0	0	1	0	0	0	1
(6)	0	0	0	1	0	1	0	0
(7)	0	0	0	1	0	0	0	0
(8)	0	0	0	1	0	1	0	1
(9)	0	1	0	0	0	0	0	1
(10)	0	1	0	0	0	1	0	0
(11)	0	1	0	0	0	0	0	0
(12)	0	1	0	0	0	1	0	1
(13)	0	1	0	1	0	1	0	1
(14)	0	1	0	1	0	0	0	0
(15)	0	1	0	1	0	0	0	0
(16)	0	1	0	1	0	1	0	1

نضع:

ζ_{im} **احتمال الحصول على المفردة** (x_0) **في الخلية** (m) **للمجموعة** $(\prod_i)$**، في الصورة الآتية:**

$$\zeta_i(Z) = \zeta_{im} \quad , \ (i = 1,2,...,n; m = 1,2,...,\ell) \qquad (1)$$

حيث :

$$(i)\ \sum_{m=1}^{t} \zeta_{im} = 1; 0 \le \zeta_{im} \le 1$$

$$(ii)\ \ell = 2^{g(k_i - 1)}$$

حيث يفترض النموذج المقترح قاعدة تصنيف وتوزيع المفردات في المجموعات المختلفة، إن التوزيع الشرطي لـ (x) **في خلية كثير الحدود** (m)**، هو** $\zeta_i(X/Z)$ **ويكون في الصورة الآتية:** (lime, w,L, 1995)

$$\zeta_i(X / Z) = \prod_{j=1}^{t} \frac{n'_{ij}!}{x_{ij}!} \phi_{imx_j}^{x_{imj}}$$

حيث :

$$(i)\ 0 \le \phi_{imx_j} \le 1. \text{is the parameter of } (x_j) \text{in cell } (m) \text{ for population } (\prod_i).$$
$$(ii)\ i = 1,2,...,m\ ;\ j = 1,2,...t\ ;\ m = 1,2,...,\ell$$
$$(iii)\ x_j = 1,2,...,n'_j$$

أما بالنسبة للتوزيع الشرطي لـ (Y/X,Z)**، فإنه يتبع التوزيع الطبيعي متعـدد المتغيـرات بالمعـالم،** $(\mu_{imx_j}, \Sigma_{mx_j})$ **بمعنى أن :**

$$f(Y / X, Z) \sim N\ (\mu_{imx_j}, \Sigma_{mx_j}) \qquad (3)$$

ويمكن التعبير عن احتمال بأن المشاهدة من المجموعة (II_i)**، في الصورة الآتية:**

$$q(\prod_i) = \varphi_i,\ i = 1,2,..., w \text{ and } \sum_{i=1}^{w} \varphi_i = 1$$

وبناء على ما سبق يمكن تحديد قاعدة توزيع المفردات على المجموعات المختلفة باستخدام نظريـة بايز، حيث تكمن المشكلة في توصيف المشاهدة x_0=(Z,X,Y) **في إحدى المجموعـات** (II, i=1,2,…,w)**.**
فإذا كانت المفردة (x_0) **من المجموعة** $(\prod_i)$**، فإنه يمكن كتابة دالة كثافة الاحتمـال فـي الصـورة الآتية:**

$$f_i(x_0) = f_i\ (Z) f_i\ (X/Z) f_i\ (Y/X,Z)$$
$$= \prod_{j=1}^{t} \alpha_i \frac{n'_{ij}!}{x_{im_j}!} \zeta_{im} \varphi_i \phi_{imx_j}^{x_{imj}} \exp(\mu_{imx_j}, \Sigma_{imx_j}) \qquad (5)$$
$$\{i = 1,2,...,w;\ j = 1,2,....,t;\ m = 1,2,...,\ell$$
$$and\ \ x_j = 0,1,2,...n'_j$$

حيث :

α_1 = مقدار ثابت، يتم اختباره ليكون مجموع الاحتمال يساوي الواحد .

ومن ثم فإن احتمال ($q(\prod_i / x_0)$) يكون في الصورة الآتية :

$$q(\prod{}_i / x_0) = \frac{q(\prod_i) q(x_0 / \prod_i)}{\sum_{i=1}^{w} q(\prod_i) q(x_0 / \prod_i)} \qquad (6)$$

وباستخدام العلاقتين (4) و (5) يتم الحصول على المقدار الآتي :

$$q(\prod{}_i / x_0) = \frac{\prod_{j=1}^{t} \alpha_i \beta_{ij} \varphi_i \zeta_{im} \phi_{imx_j}^{x_{imj}} ex[\{-\frac{1}{2}(Y-\mu_{imx_j})' \sum_{m}^{n} (Y-\mu_{imx_j})}{\sum_{i=1}^{w} \prod_{j=1}^{t} \alpha_i \beta_{ij} \varphi_i \zeta_{im} \phi_{imx_j}^{x_{imj}} ex[\{-\frac{1}{2}(Y-\mu_{imx_j})' \sum_{m_j}^{n} (Y-\mu_{imx_j})} \quad (7)$$

حيث : $\beta_{ij} = \dfrac{n'_{ij}!}{X_{im_j}!};\quad i = 1,2,...,w; m = 1,2,...,\ell\ and\ X_j = 0,1,2,...,n'_{ij}$

وتوضع المقادير :

$$\beta_{ivmx_j} = \sum\nolimits_{m_j}^{n} (\mu_{imx_j} - \mu_{vmx_j}),$$
$$C_{ivmx_j} = -\frac{1}{2}(\mu_{imx\,j} - \mu_{imx_j})' \sum\nolimits_{m}^{n} (\mu_{imx_j} + \mu_{vmx_j}$$
$$\gamma_{ivmx_j} = ln \frac{\alpha_i \beta_{ij} \varphi_i \zeta_{im} \phi_{imx_j}^{X_{imj}}}{\alpha_v \beta_{vj} \varphi_v \zeta_{vm} \phi_{vmx_j}^{X_{vmj}}}$$

فإنه يمكن كتابة العلاقة (7) في الصورة الآتية :

$$q(\prod_i / x_0) = \frac{\prod_{j=1}^{t} \exp\{Y'B_{ivmx_j} + C_{ivmx_j} + \gamma_{ivmx\,j}\}}{1 + \sum \exp \prod_{j=1}^{t} \{Y'B_{ivm_j} + C_{ivmx_j} + \gamma_{ivmx\,j}\}} \qquad (8)$$

و:

$$q(\prod_i / x_0) = \frac{\exp \sum_{j=1}^{t} \{Y'B_{ivmx_j} + C_{ivmx_j} + \gamma_{ivmx\,j}\}}{1 + \sum_{i \neq v}^{w} \exp \sum_{j=1}^{t} \{Y'B_{ivm_j} + C_{ivmx_j} + \gamma_{ivmx\,j}\}} \qquad (9)$$

وبوضع المقدار :

$$\sum_{j=1}^{t} \{Y'B_{ivmx_j} + C_{ivmx_j} + \gamma_{vmx_j}\} = n_{imx_j}$$

فإنه يمكن إعادة كتابة العلاقة (8) **في الصورة الآتية :**

$$q(\prod_i / x_0) = \frac{e^{n_{imx\,j}}}{1 + \sum_{i=1}^{w} e^{n_{imx\,j}}} \qquad (10)$$

وتكون قاعدة تصنيف المفردات هي :

Allocate an observation x_0 (Z,X,Y) into population ($\prod_i$) if:

$$(i) \quad n_{imx_j} \geq 0$$

$$(ii) \; n_{imx_j} < 0 \qquad (11)$$

وفي التطبيقات العملية لقواعد تصنيف وتوزيع المفردات في المجتمعـات المختلفـة، قـد يتضـمن التصنيف بعض الأخطاء، بمعنى أنه قد تصنف المفردات في مجتمع غير المجتمع الذي تنتمي إليـه، ومن ثم يتطلب تقييم لمدى كفاءة وسلامة هذه القواعد في تصنيف المفردات - ويتم في الغالب تقييم هذه القواعد بدلالة احتمالات أخطاء التصنيف.

وطبقاً للنموذج المقترح ـ المذكور سابقاً ـ فإن إجمالي احتمال التصنيف الخطأ Total Probability of misclassification **في المجتمع ($\prod_i$)، هو عبارة عن مجموع احتمال التصنيف الخطأ في كل خلية من كثير الحدود في المجتمع ($\prod_i$) مرجحاً باحتمال** (X_i) **في هذه الخلية.**

وبوضع المقدار :

$$A = \left\{ \frac{\theta_i c(v/i)}{\theta_v c(i/v)} \right\}$$

فإن احتمال التصنيف الخطأ Probability of misclassification **للمفردة** (x_0) **في المجتمع ($\prod_v$) بينما هي في الحقيقة تنتمي إلى المجتمع** (IIi)**، يكون في الصورة الآتية:**

$$q(v/i) = \sum_{m}^{t} \sum_{x}^{n'} \prod_{j=1}^{p} \frac{n'_{im}!}{X_{imj}!} \zeta_{im} \phi_{imx_j}^{imj} F\{(Log\ A - \frac{1}{2} D_{mx_j}^2) / D_{mx_j}\} \qquad (13)$$

كما أن احتمال التصنيف الخطأ Probability of misclassification **للمفردة** (x_0) **في المجتمع** (II_i) **في حين هي في الحقيقة تنتمي إلى المجتمع** (II_v)**، يكون في الصورة الآتية:**

$$q(v/i) = \sum_{m}^{t} \sum_{x}^{n'} \prod_{j=1}^{p} \frac{n'_{im}!}{X_{imj}!} \zeta_{im} \phi_{imx_j}^{imj} F\{(Log\ A - \frac{1}{2} D_{mx_j}^2) / D_{mx_j}\} \qquad (14)$$

حيث :

(i) F = **تابع التوزيع الطبيعي متعدد المتغيرات**

$$(ii) D_{mx_j}^2 = (\mu_{imx_j} - \mu_{vmx_j})' \sum_{m_j}^{n} (\mu_{imx_j} - \mu_{vmx_j}) (\prod_i, \prod_v)$$

تصنيف المفردات بين مجتمعين

(iii) c (v / i) = $\prod_v$ **تكلفة التصنيف الخطأ لمفردات المجتمع في المجتمع**

4 ـ تقدير معدلات الخطأ لأسلوب التحليل التميزي:

في كثير من التطبيقات العملية لأسلوب التحليل التمييزي للفصل بين المجتمعات، قد يتضمن التحليـل أخطاء في تصنيف المشاهدات، بمعنى أنه قد تصنف المشاهدات في مجتمع غير المجتمع الذي تنتمي إليه. وعليه فإنه بعد أن يتم إيجاد وتحديد دوال التمييز للفصل بين المجتمعات، يتطلب إيجاد مقيـاس

للحكم على مدى كفاءة هذه الدوال والاعتماد عليها لتصنيف المشاهدات في تلك المجتمعات في الدراسات المستقبلية. (Ahmed, M,S.1998)

إن أحد الأساليب المهمة لتقييم مدى كفاءة دوال التمييز للفصل بين المجتمعات وتوصيف المشاهدات في هذه المجتمعات هو إيجاد معدلات الخطأ Estimation of error rates **لتصنيف المشاهدات، أو ما يطلق عليه احتمال التصنيف الخطأ للمشاهدات** Misclassifcation probability **ويفرض أن** $f_1(x)$ ، $f_2(x)$ **هما دالتي كثافة الاحتمال للمجتمعين** $\prod_1$ ، $\prod_2$ **على الترتيب، وإذا كانت :**

أ - المعالم (μ_1, μ_2, V) **معلومة.**

حيث إن :

η_1 **: متوسط مفردات المجتمع الأول.**

η_1 **: متوسط مفردات المجتمع الثاني.**

V **= مصفوفة التغايرات في المجتمعين .**

ب - الاحتمالات القبلية للمجتمعين هي q_1 **للمجتمع (** $\prod_1$ **)،** q_2 **للمجتمع (** $\prod_2$ **)، حيث** $q_1+q_2=1$ **.**

فإن احتمالات التصنيف الخطأ للمشاهدات تكون على النحو الآتي:

1 **- احتمال أن تكون المشاهدة من المجتمع الأول (** $\prod_1$ **) وصنفت خطأ بأنها من المجتمع الثاني (** $\prod_2$ **)، يكون في الصورة الآتية :**

$$q(2/1) = q(x \in R_2 \mid \textstyle\prod_1) = \int_{R_2} f_1(x)dx \qquad (15)$$

2 **- احتمال أن تكون المشاهدة من المجتمع الثاني (** $\prod_2$ **) وصنفت خطأ بأنها من المجتمع الأول (** $\prod_1$ **)، يكون في الصورة الآتية :**

$$q(1/2) = q(x \in R_1 \mid \textstyle\prod_2) = \int_{R_1} f_2(x)dx \qquad (16)$$

حيث :

R_1**: المنطقة التي تصنف فيها مشاهدات المجتمع الأول (** $\prod_1$ **).**

R_2 **: المنطقة التي تصنف فيها مشاهدات المجتمع الثاني (** $\prod_2$ **).**

ويمكن اشتقاق احتمال التصنيف الخطأ للمشاهدات كحاصل ضرب الاحتمالات القبلية والشرطية على النحو التالي :

1 ـ **احتمال أن تكون المشاهدة من المجتمع الأول ($\prod_1$) وصنفت خطأ بأنها من المجتمع الثاني ($\prod_2$) يكون في الصورة الآتية :**

$$q(x \in R_2 \mid \prod{}_1)q(\prod{}_1) = q(2/1)q_1 \qquad (17)$$

2 ـ **احتمال أن تكون المشاهدة من المجتمع الثاني ($\prod_2$) وصنفت خطأ بأنها من المجتمع الأول ($\prod_1$) يكون في الصورة الآتية :**

$$q(x \in R_1 \mid \prod{}_2)q(\prod{}_2) = q(1/2)q_2 \qquad (18)$$

ويكون الاحتمال الكلي للتصنيف الخطأ Total probability of misclassification **للمشاهدات في الصورة الآتية :**

$$T(R,f) = q_1 \int_{R_2} f_1(x)dx + q\int_{R_1} f_2(x)dx \qquad (19)$$

وتتحقق أقل قيمة لهذا الاحتمال عندما يتم تحديد مناطق التصنيف (R_2) ، (R_1) **للمشاهدات تحديداً سليماً، حيث يطلق على هذه القيمة معدل الخطأ الأمثل** Optimum error rate **لتصنيف المشاهدات، ويكون في الصورة الآتية :**

$$OER(R,f) = q_1 \int_{R_2} f_1(x)dx + q\int_{R_1} f_2(x)dx \qquad (20)$$

ويكون المقدار (OER) **هو معدل الخطأ الأمثل لأقل احتمال كلي** (TMP) **الذي يستخدم للحكم على مدى كفاءة وسلامة تصنيف المشاهدات .**

إن طريقة التصنيف المثلى للمشاهدات يجب أن تأخذ في الحسبان التكلفة المرتبطة بأخطاء التصنيف لتلك المشاهدات. ويمكن التعبير عن تكاليف تلك الأخطاء في صورة مصفوفة التكاليف الآتية (Anderson, T.W.1984) **:**

	$\prod_1$	$\prod_2$
$\prod_1$	0	c (2/1)
$\prod_2$	c(1/2)	0

ومن ثم فإن تكاليف تصنيف المشاهدات باستخدام المصفوفة السابقة تكون في الصورة الآتية:

1 ـ صفر في حالة التصنيف الصحيح للمشاهدة .

2 ـ c(1/2) في حالة التصنيف الخطأ لمشاهدة من المجتمع الثاني ($\prod_2$) على أنها من المجتمع الأول ($\prod_1$).

3 ـ c(2/1) في حالة التصنيف الخطأ لمشاهدة المجتمع الأول ($\prod_1$)على أنها من المجتمع الثاني ($\prod_2$) .

ويمكن اشتقاق قاعدة التكاليف المتوقعة للتصنيف الخطأ Expected cost of misclassification للمشاهدات باستخدام مصفوفة التكاليف السابقة وذلك بضرب العناصر غير القطرية للمصفوفة في احتمالات حدوثها وذلك في الصورة الآتية :

ECM = c(2/1) q(2/1)q_1 + c(1/2) q (1/2)q_2 . (21)

حيث :

c(2/1) = تكلفة التصنيف الخطأ لمفردات المجتمع الأول ($\prod_1$) في المجتمع الثاني ($\prod_2$).

c(1/2) = تكلفة التصنيف الخطأ لمفردات المجتمع الثاني ($\prod_2$) في المجتمع الأول ($\prod_1$)

وبفرض أن (Ω) تشير إلى فراغ العينة الذي يتضمن جميع المشاهدات الممكنة (x)، فإنه يمكن كتابة فراغ العينة في الصورة الآتية :

$$\int_{\Omega} f_1(x)dx = \int_{R_1} f_1(x)dx + \int_{R_2} f_1(x)dx = 1, \quad \Omega = R_1 \cup R_2 = 1 \qquad (22)$$

بشرط $R_1 \cap R_2 \neq \phi$

وباستخدام العلاقة (22) وقاعدة أدنى احتمال كلي للتصنيف الخطأ

$$\hat{Q}_k(x) > \{\hat{Q}_1(x), \hat{Q}_2(x), \ldots\ldots, \hat{Q}_g(x)\}$$

يمكن كتابة العلاقة السابقة بالصورة الآتية :

$$ECM = c(2\backslash 1)q_1\left[1 - \int_{R_1} f_1(x)dx\right] + c(1/2)q_2 \int_{R_1} f(x)dx$$

$$ECM = \int_{R_1} [c(1/2)q_2 f_2(x) - c(2/1)q_1 f_1(x)]dx + c(2/1)q_1 \qquad (23)$$

وحيث إن المقادير (q_1, q_2) **،** $[c(2/1), c(1/2)]$ **،** $f_2(x), f_1(x)$ **قيم غير سالبة لجميع قيم** (x) **، فإنه يتم تصغير** (ECM) **إذا تضمنت المنطقة** (R_1) **قيم** (x) **ويكون عندها المقدار:**

$$[c(1/2)\, q_2 f_2(x) - c(2/1) q_1 f_1(x)] \le 0$$

ومن ثم فإن منطقتي التصنيف الأمثل $(R_2), (R_1)$ **اللتين تحققان أقل متوسط تكلفة للتصنيف الخطأ يمكن تحديدها باستخدام المتباينة الآتية :**

$$c(1/2)\, q_2\, f_2(x) \le c(2/1)\, q_1\, f_1(x)$$

$$R_1 : \frac{f_1(x)}{f_2(x)} \ge \left[\frac{c(1/2)}{c(2/1)}\right]\left[\frac{q_2}{q_1}\right] \qquad (25)$$

$$\begin{pmatrix} density \\ ratio \end{pmatrix} \ge \begin{pmatrix} \cos t \\ ratio \end{pmatrix}\begin{pmatrix} prior \\ prob.ratio \end{pmatrix}$$

$$R_2 : \frac{f_1(x)}{f_2(x)} < \left[\frac{c(1/2)}{c(2/1)}\right]\left[\frac{q_2}{q_1}\right]$$

وسوف يتم تحديد مناطق التصنيف وذلك إذا فرضنا أن :

1 **ـ الاحتمالات القبلية** $(q_2 / q_1=1)$ **متساوية فإن منطقتي التصنيف للمجتمع الأول** (II_1) **والمجتمع الثاني** (II_2) **اللتين تحققان أقل متوسط تكلفة للتصنيف الخطأ يتم تحديدهما في الصورة الآتية:**

$$R_1 : \frac{f_1(x)}{f_2(x)} \ge \frac{c(1/2)}{c(2/1)};\ R_2 : \frac{f_1(x)}{f_2(x)} < \frac{c(1/2)}{c(2/1)} \qquad (26)$$

2 **ـ تكاليف التصنيف الخطأ** $c(1/2) / c(2/1)=1$ **متساوية ، فإن منطقتي التصنيف** (R_2, R_1) **اللتين تحققان أقل متوسط تكلفة للتصنيف الخطأ يتم تحديدهما كالآتي:**

$$R_1 : \frac{f_1(x)}{f_2(x)} \ge \frac{q_2}{q_1};\ R_2 : \frac{f_1(x)}{f_2(x)} < \frac{q_2}{q_1} \qquad (27)$$

3 ـ الاحتمالات القبلية وتكاليف التصنيف الخطأ $(\frac{q_2}{q_1} = c(1/2)\ /\ c(2/1) = 1)$ متسـاوية ، فإن منطقتي التصنيف اللتين تحققان أقل متوسط تكلفة للتصنيف الخطأ يتم تحديدهما كالآتي:

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq 1; \qquad R_2 : \frac{f_1(x)}{f_2(x)} < 1 \qquad (28)$$

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq 1; \qquad R_2 : \frac{f_1(x)}{f_2(x)} < 1$$

ويمكن اشتقاق قاعدة تصنيف المشاهدات في حالة مجتمعين يتبعان التوزيع الطبيعي متعدد المتغيرات على النحو الآتي :

بفرض أن دالة الكثافة المشتركة للمجتمعين $(\prod_1)$،$(\prod_2)$ هي $f_i(x)$ عند المتجه $[X_1,X_2,\ldots X_p]$ = X^x، وأن (μ_1, μ_2) معلومة، ومصفوفتي التغاير $\Sigma_1=\Sigma_2=\Sigma$ للمجتمعين متساوية. فإن مناطق التصنيف (R_2, R_1) في العلاقة (28) اللتين تحققان أدنى متوسط تكلفة متوقعة للتصـنيف الخطـأ (ECM) يمكـن تحديدهما في الصورة الآتية :

$$R_1 : exp\left[-\frac{1}{2}(x-\mu_1)^1 \sum (x-\mu_1) + -\frac{1}{2}(x-\mu_2)^1 \sum (x-\mu_2)\right] \geq \left(\frac{c(1/2)}{c(2/1)}\right)\left(\frac{q_2}{q_1}\right)$$

$$R_2 : exp\left[-\frac{1}{2}(x-\mu_1)^1 \sum (x-\mu_1) + -\frac{1}{2}(x-\mu_2)^1 \sum (x-\mu_2)\right] < \left(\frac{c(1/2)}{c(2/1)}\right)\left(\frac{q_2}{q_1}\right)$$

ولما كانت المقادير في العلاقة (28) غير سالبة لجميع المتجهات (x)، أمكن أخذ لوغاريتماتها الطبيعية دون تغيير اتجاه المتباينتين، ومن ثم فإن قاعدة التصنيف التي تحقق أدنى متوسـط تكلفـة متوقعـة للتصنيف الخطأ (ECM) تكون في الصورة الآتية :

$$1 - Allocate\ x_0\ to\ population\ (\prod_1)\ if\ :$$

$$(\mu_1 - \mu_2)'\,\Sigma^{-1} x_0 - \frac{1}{2}(\mu_1 - \mu_2)'\,\Sigma^{-}(\mu_1 - \mu_2) < \ln\left[\left(\frac{c(1/2)}{c(2/1)}\right)\left(\frac{q_2}{q_1}\right)\right] \qquad (29)$$

$$2 - Allocate\ x_0\ to\ population\ (\prod_1)\ if\ :$$

$$(\mu_1-\mu_2)'\,\Sigma^{-1}x_0-\frac{1}{2}(\mu_1-\mu_2)'\,\Sigma^{-}(\mu_1-\mu_2)<\ln\left[\left(\frac{c(1/2)}{c(2/1)}\right)\left(\frac{q_2}{q_1}\right)\right]$$

وإذا كانت المعالم (μ_1, μ_2, Σ) **غير معلومة، فإن قاعدة التصنيف المقدرة التي تحقق أقل متوسط تكلفة للتصنيف الخطأ** (ECM) **للمشاهدات تكون في الصورة الآتية:**

$$1-Allocate\ x_0\ to\ population\ (\prod{}_1)\ if\ :$$

$$(\bar{x}_1-\bar{x}_2)S^{-1}_{pooled}x_0-\frac{1}{2}(\bar{x}_1-\bar{x}_2)'S^{-1}_{pooled}(\bar{x}_1-\bar{x}_2)'\geq 1n\left[\left(\frac{c(1/2)}{c(2/1)}\right)\left(\frac{q_2}{q_1}\right)\right] \qquad (30)$$

$$2-Allocate\ x_0\ to\ population\ (\prod{}_2)\ if\ :$$

$$(\bar{x}_1-\bar{x}_2)S^{-1}_{pooled}x_0-\frac{1}{2}(\bar{x}_1-\bar{x}_2)'S^{-1}_{pooled}(\bar{x}_1-\bar{x}_2)'\geq 1n\left[\left(\frac{c(1/2)}{c(2/1)}\right)\left(\frac{q_2}{q_1}\right)\right]$$

5 ـ اختبار معنوية الفروق بين المجتمعات :

بعد أن تم عرض ومناقشة أسلوب التحليل التمييزي للفصل بين المجتمعات وتحديد القواعد المرتبطة بتخصيص المشاهدات على هذه المجتمعات في الدراسات المستقبلية. فإنه يتطلب اختبار الفروق بين متجهات متوسطات تلك المجتمعات، بمعنى هل هناك فروق معنوية بين هذه المجتمعات محل البحـث والدراسة؟. وفي هذا الجانب يتم استخدام التوزيع الإحصائي (F) **لإجراء مثل هذا الاختبار.**

بفرض أن المجتمعين ($\prod_1$ **)،(** $\prod_2$ **) مجتمعان معتدلان متعددا المتغيرات، وأن لهما مصـفوفة التباينات والاختلافات** ($V_1 = V_2 = V$)**، والمتوسطات** ($\mu_2 . \mu_1$) **فإذا تـم سـحب عينـة عشـوائية مـن المجتمعين. فإن الاختبار الإحصائي لمعنوية الفروق يكون في الصـورة الآتيـة** (Lachenbruch, P.A, 1975) **:**

$H_0 = \mu_1 = \mu_2$
$H_1 = \mu_2 \neq \mu_2$

$$F=\left(\frac{n_1n_2}{n_1+n_2}\right)\left(\frac{n_1+n_2-p-1}{(n_1+n_2-2)p}\right)D^2$$

Where:
d.f. = ($v_1 = p$) , $v_2 = (n_1 + n_2 - p - 1)$ (31)

$$D^2 = (\bar{x}_1 - \bar{x}_2)^1 S^{-1} (\bar{x}_1 - \bar{x}_2)$$

6 ـ الخاتمة:

تطرقنا في هذا البحث إلى المستويات الاقتصادية والاجتماعية المختلفة والتي تشكل فيما بينها الهيكل البنائي للنسيج الاقتصادي والاجتماعي لأسر المجتمع، كما بينا كيفية توصيف وتوزيع مفردات الدراسة داخل هذه المستويات. وفي هذا الجانب تُستَخدم دوال التمييز لتوصيف وتوزيع مفردات الدراسة داخل هذه المستويات وتحديد المستوى الاقتصادي والاجتماعي الذي تنتمي إليه المفردة سواء كانت خارج نطاق هذا الهيكل أو التي تنتقل من مستوى اقتصادي اجتماعي إلى مستوى اقتصادي اجتماعي آخر داخل هذا الهيكل. وبينّا أيضاً مدى كفاءة دوال التمييز لتوصيف وتوزيع الأسر من خلال تقدير معدلات الخطأ لنصل أخيراً إلى اختبار معنوية الفروق بين المجتمعات للتأكد من معنوية تصنيف وتوزيع الأسر.

نأمل من الباحثين في مجال العلوم الاقتصادية والاجتماعية وعلاوة على تحديد الهيكل الاقتصادي الاجتماعي لأسر المجتمع وتوصيف وتوزيع مفردات الدراسة داخل هذه المستويات، الاهتمام بدراسة أبعاد ومتغيرات اقتصادية واجتماعية متجددة، وتحديد العوامل المؤثرة في تحديد الهيكل الاقتصادي الاجتماعي لأسر المجتمع وتوصيف وتوزيع الأسر داخل مستويات هذا الهيكل، و ضرورة صياغة نموذج التوصيف والتوزيع الأمثل لأسر المجتمع المدروس داخل مستويات الهيكل الاقتصادي الاجتماعي.

أخيراً إن الأفكار التي قدمت في هذا البحث تفتح المجال لدراسات لاحقة، تهدف إلى:

- **تحديد هيكل اقتصادي اجتماعي أمثل لأسر المجتمع، بهدف الوصول إلى قاعدة معلوماتية توفر أرضية يمكن أن تبنى عليها قرارات إجرائية وتنفيذية لدى المخططين والمنفذين في أجهزة الدولة بما يحقق الاستقرار الاقتصادي والاجتماعي للفرد والأسرة في المجتمع.**
- **إعداد دراسات أخرى تسهم في وضع سياسات اقتصادية ـ اجتماعية تهدف إلى العدالة في في توزيع الدخل والعبء الضريبي والإعانات الحكومية لأسر المجتمع بصورة أكثر واقعية.**

التوصيات :

بناء على ما تقدم، وفي ضوء الدراسة، فإن الباحث يقترح التوصيات الآتية:

1 ـ يجب على المخططين والمنفذين في أجهزة الدولة الاهتمام والتركيز على الأسر ذات المستوى الاقتصادي والاجتماعي المنخفض، وتحسين أوضاعهم الاقتصادية والاجتماعية لتحقيق الاستقرار الاقتصادي والاجتماعي للفرد والأسرة في المجتمع.

2 ـ الاهتمام بعملية توصيف وتوزيع أسر المجتمع داخل الهيكل البنائي الاقتصادي والاجتماعي من خلال استخدام أسلوب التحليل التمييزي للمستويات الاقتصادية والاجتماعية التي يتكون منها الهيكل البنائي لمجتمع الدراسة، سواء كانت أسر المجتمع خارج نطاق هذا الهيكل أو التي تنتقل من مستوى اقتصادي واجتماعي إلى مستوى اقتصادي واجتماعي آخر داخل هذا الهيكل.

3 ـ إجراء دراسات مستمرة على الأسرة السورية بإدخال مزيد من المتغيرات الاقتصادية والاجتماعية، التي يمكن أن تؤثر في تحديد الهيكل البنائي للنسيج الاقتصادي والاجتماعي لأسر المجتمع السوري.

المراجع

المراجع العربية :

1 ـ عبد الفتاح محمد قنديل : الإحصاء في العلوم الاجتماعية، دار وهدان للطباعة والنشر، القاهرة، 1995.

2 ـ عبد المرضي عزام، التحليل الإحصائي للمتغيرات المتعددة، دار المريخ للطباعة والنشر، الرياض 1998.

المراجع الأجنبية:

1- Anderson, T.W. (1984) " An Introduction to Multivariate Statistical. Analysis" John Wiley & Stons, Inc, New York.

2- Brynaf, J.Manly, (1994) "Multivariate statistical Methods", Chapman & Hall, New York.

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

4- Kandil, A.M., (1992) " Discriminant with mixtures of continuous, discrete and nominal variables", The Egyptian Journal, ISSR, Cairo Univ. vol. 36. No 1, 102, 117.

5- Krzanowski, W.J. (1995). "Multivarite Analysis Classification, Covariance structures", John wiley * Stons, Inc, New York.

6- Lachenbruch, P.A. (1985). "Discriminant Analysis" Macmillan Publishing Co, Inc.

7- Romeburg, H.C, (1984) "Cluster Analysis for Reseacher" Lifetime learing publication, New York.

8- Lime, W.L, (1995)" On linear Discriminat Analysis with Adaptive Redge Classication Rules", J.Multi. Ana, 53, 264-290.

الرسائل العلمية :

1- Ahmed, M.S., (1998), "A comparision of the Discriminant and logistic Regression Approaches", Ph.D thesis, IssR, Cario University.

2- Gouda, Mohamed Khalil, (1999), A study to some methods which are used for inducting the probability distributions from data about probability phenomena", M.Sc. thesis for degree in statisties, Faculity of Comm, Zagazig University.*

* تاريخ ورود البحث إلى مجلة جامعة دمشق 2006/8/16.