

# الأسس الرياضية للتحليل التمييزي (الخطي والنوعي)

## الأسس الرياضية للتحليل اللوجستي والتحليل العنقودي

الدكتور إبراهيم محمد العلي (2019)

استاذ في كلية الاقتصاد بجامعة تشرين - سورية

أتقدم بهذا المنشور العلمي إلى جميع الباحثين العرب آملاً أن يستفيدوا منه في أبحاثهم العلمية المختلفة، مع رجاء الإشارة إلى ذلك .

وهذا المنشور يتألف من سبعة فصول وملحقين تشمل ما يلي:

الأسس الرياضية للتحليل التمييزي (الخطي والنوعي): الفصول (1) و(2) و(3) و(4) و(5) .

الأسس الرياضية للتحليل اللوجستي: الفصل (6) .

الأسس الرياضية للتحليل العنقودي: الفصل (7) .

الملحق (1): عن المصفوفات وخواصها والعمليات عليها .

الملحق (2): عن القيم الذاتية (الكامنة) للمصفوفات الشاذة .

ولقد اعتمدت عند إعداد هذا المنشور على المراجع العلمية والمنشورات الالكترونية المذكورة في نهاية المنشور .

وأغتنم هذه المناسبة لأتقدم بجزيل الشكر والامتنان للزميلين الدكتور طالب أحمد والدكتور ياسر علوش على مساعدتهما لي في المراجعة والتدقيق، كما أشكر الأنسة سوزان صقر على تحملها أعباء الطباعة والتفتيح، وعلى كل ما بذلته من جهد وفنٍ في إخراج هذا المنشور إلى الفضاء الالكتروني . واعتذر شخصياً للقراء عن كل خطأ أو سهوٍ لم أستطع تداركه خلال تلك العمليات .

مع أطيب التمنيات

2019/2/12

أ.د. إبراهيم محمد العلي

# الأسس الرياضية للتحليل التمييزي (الخطي والنوعي)

الدكتور إبراهيم محمد العلي

استاذ في كلية الاقتصاد بجامعة تشرين - سورية

## الفصل الأول

### المفاهيم العامة للتحليل التمييزي

#### 1-1 : تمهيد:

إن كل إنسان يمارس التحليل التمييزي يومياً وبصورة لا شعورية . فعندما نسير في الشارع نميز فوراً بين الناس حسب الجنس ونفرزهم إلى ذكور وإناث، كما نميز بينهم حسب العمر ونفرزهم دون أن نسألهم إلى كبار وشباب وأطفال . والطبيب عندما يفحص المرضى يميز بينهم بسرعة حسب حالاتهم الصحية (مريض أو سليم)، وذلك قبل إجراء التحاليل الطبية اللازمة، وأحياناً يعالجهم دون إدخالهم المستشفى. والمتجول في الغابة يميز بين أنواع الأشجار بسرعة، والشرطي الذي يقف على بوابة المطار الدولي يميز في لحظة خاطفة بين المسافرين حسب جنسياتهم ويفرزهم إلى أوروبيين وأفارقة وهنود وصينيين ... الخ . وهكذا نجد أن عمليات التمييز كثيرة جداً وهي تهدف إلى تحديد المجموعة التي ينتمي إليها أي عنصر من عناصر المجتمع المدروس أو تحديد انتماء أي عنصر جديد إلى إحدى مجموعات المجتمع المدروس، ولكن عملية التمييز بين العناصر وتحديد المجموعات التي تنتمي إليها ليست بهذه البساطة والسهولة .

فمثلاً لا نستطيع ببساطة التمييز بين الأخيار والأشرار أو بين المثقفين وغير المثقفين أو بين الأذكاء والأغبياء ... الخ، وذلك لأن التمييز في مثل هذه الحالات يحتاج إلى دراسة معمقة وتحديد المؤشرات التي تساعدنا في عملية التمييز .

وهنا نطرح الأسئلة التالية :

1- كيف استطعنا فرز هؤلاء الناس وتحديد المجموعات التي ينتمون إليها وبهذه السرعة الفائقة ؟

2- ماهي المتحولات الكمية أو النوعية التي اعتمدنا عليها في عمليات الفرز وتحديد الانتماء ؟

3- ماهي الصيغ أو التراكيب الرياضية التي استخدمناها وطبقناها على تلك المتحولات لاتخاذ القرار

المناسب لفرز هذه العناصر وتحديد المجموعات التي تنتمي إليها ؟

إن الإجابة على هذه الأسئلة هي: إننا استخدمنا بعض الصفات أو بعض الخواص الملازمة لهذه العناصر وقمنا بتركيبها ضمن صيغ معينة ( فورية )، ومنها حصلنا على مؤشر معين ساعدنا على فرز تلك العناصر وعلى تحديد المجموعات التي تنسب إليها .

ولكن بعض عمليات الفرز والتمييز قد لا تكون صحيحة بسبب التشابه بين خواص العناصر، أو بسبب الالتباس بينها . ففي مثال شرطي المطار الدولي فقد يقع في خطأ ما . فالشخص الذي اعتبره اوروبياً قد يكون كندياً والإفريقي قد يكون أمريكياً والهندي قد يكون باكستانياً والصيني قد يكون يابانياً ... الخ . وهكذا نجد أن عمليات التمييز بين العناصر ليست أمراً سهلاً، وهي تحتاج إلى دراسات معمقة وإلى أدوات رياضية وإحصائية متقدمة، يأتي في مقدمة هذه الأدوات ما يسمى التابع التمييزي، الذي يستخدم لفرز عناصر العينة المدروسة إلى المجموعات المناسبة لها ( مع احتمال الخطأ في ذلك ) .

وقبل أن ندخل في استعراض التحليل التمييزي نشير إلى بعض المصطلحات المتشابهة والمتداخلة، وذلك حتى يتم التفريق بينهما عند معالجة المسائل المختلفة وهذه المصطلحات هي [حسب [4] ص437]

1- التمييز: وهو يشترط وجود مجتمع مقسم إلى مجموعتين أو أكثر (محددتين مسبقاً)، وأن يتم سحب عينة عشوائية من كل منها، ويتلخص هدفه ومهمته في دراسة خواص هذه العينات، ثم استخلاص قاعدة معينة مبنية على بيانات العينة، تسمح لنا بفرز عناصر المجتمع إلى تلك المجموعات وتحديد الحدود الفاصلة بينها، وبحيث تمكننا من تحديد المجموعة التي ينتمي إليها أي عنصر جديد، دون أن يكون لدينا أي فكرة مسبقة عن المجموعة ينتمي إليها ذلك العنصر .

2- التصنيف: وهو يتناول توزيع عناصر المجتمع أو عينة منه إلى مجموعات محددة، وبحيث يتم تصنيف عناصر تلك العينة (أو المجتمع) إلى مجموعات متجانسة داخلياً ومختلفة عن بعضها البعض أكثر ما يمكن، وينتج عنه مجموعات محددة وذات مواصفات معينة.

أي أن التمييز يشترط مسبقاً وجود المجموعات (المجموعات) المحددة، ولكن التصنيف يهدف إلى تشكيل هذه المجموعات المختلفة في المجتمع .

3- التجزئة: وهي تتناول تجزئة عناصر عينة معينة (أو عناصر المجتمع الكلي)، وتقسيمها إلى مجموعات جزئية، بغض النظر عن وجود حدود طبيعية لتلك التجزئة أم لا، مثل توزيع الطلاب عشوائياً على الشعب الدراسية بدون أية شروط مسبقة .

وهنا نشير إلى أن التحليل التمييزي يتناول عمليتين أساسيتين هما:

التمييز: وهو القيام بفرز عناصر المجتمع إلى مجموعات محددة ومنفصلة وتحديد الحدود الفاصلة بينها .  
التصنيف: وهو القيام بالتنبؤ بانتماء أي عنصر جديد إلى إحدى المجموعات وفق قاعدة معينة .

## 1-2 تعريف التحليل التمييزي:

إن التحليل التمييزي هو عبارة عن تقنية إحصائية رياضية قوية، تستخدم لتوصيف عناصر المجتمع المدروس، والموزعة على مجموعات محددة ومنفصلة ومتكاملة (أكثر أو يساوي من مجموعتين)، وتحديد الحدود الفاصلة بينها، واستخلاص قاعدة معينة لتحديد انتماء أي عنصر إليها، ونرمز لهذه المجموعات بـ  $G_1, G_2, G_3 \dots G_g$ ، حيث أن:  $g$  هو عدد تلك المجموعات في المجتمع المدروس، وأن  $(g \geq 2)$  .

كما يستخدم التحليل التمييزي للتنبؤ بانتماء أي عنصر جديد إلى إحدى المجموعات السابقة أو تصنيفه في احداها حسب قاعدة معينة .

لذلك يقسم التحليل التمييزي إلى قسمين أساسيين هما:

- التحليل التمييزي الوصفي: وهو الذي يتناول توصيف العناصر ضمن المجموعات ووضع قواعد الانتماء إليها .

- التحليل التمييزي التنبؤي: وهو الذي يتناول تصنيف العناصر إلى مجموعات ويتنبأ بانتماء أي عنصر جديد إلى إحدى تلك المجموعات .

كما يمكن تقسيم التحليل التمييزي إلى نوعين آخرين هما:

- التحليل التمييزي البسيط: وهو الذي يعتمد على متحول واحد  $X$  لتوصيف عناصر المجتمع وتصنيفها ضمن مجموعاته المتعددة ( $g \geq 2$ ) .

- التحليل التمييزي المتعدد: وهو الذي يعتمد على عدة متحولات مستقلة (متحولين أو أكثر) لتوصيف عناصر المجتمع وتصنيفها ضمن مجموعاته المتعددة ( $g \geq 2$ ). ونرمز لهذه المتحولات بالرموز  $X_1, X_2, \dots, X_p$  حيث أن  $p$  هو عدد المتحولات المستقلة .

وكذلك يمكن تقسيم التحليل التمييزي إلى نوعين آخرين هما:

- التحليل التمييزي الخطي: وهو الذي يعتمد على النماذج الخطية للفصل بين مجموعات المجتمع المدروس .

- التحليل التمييزي غير الخطي: هو الذي يعتمد على النماذج غير الخطية للفصل بين مجموعات المجتمع المدروس .

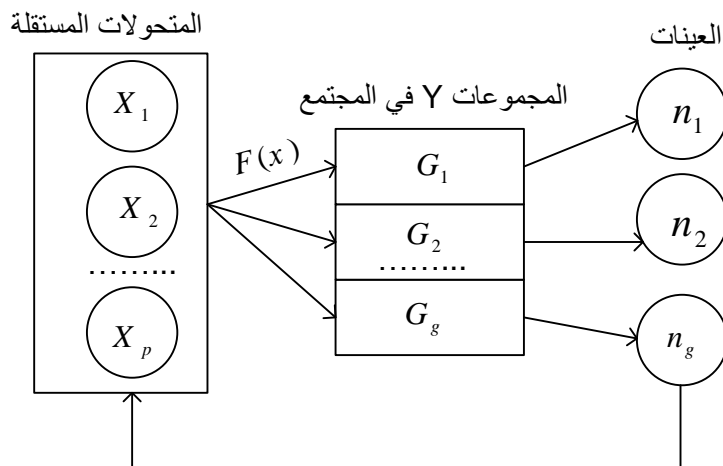
وأخيراً يمكن تقسيم التحليل التمييزي إلى نوعين آخرين هما:

- التحليل التمييزي الكمي: وهو الذي يعتمد على المتحولات الكمية (المقاسة) في توصيف عناصر المجتمع، (كالدخل والوزن والعمر... الخ) .

- التحليل التمييزي النوعي: وهو الذي يعتمد على المتحولات النوعية (غير المقاسة) في توصيف عناصر المجتمع (كالجنس والتعليم والمكان والحالة الاجتماعية... الخ) .

وسنركز في منشورنا هذا على التحليل التمييزي الكمي الخطي الذي يهدف إلى دراسة التأثير المشترك لعدد متحولات كمية مستقلة هي  $X_1, X_2, \dots, X_p$  على تابع نوعي  $Y$  مؤلف من عدة مجموعات (أو فئات) هي:  $G_1, G_2, \dots, G_g$  وتحديد دور كل منها في تصنيف عناصر المجتمع إلى هذه المجموعات ، وذلك من خلال بيانات العينات العشوائية المسحوبة من تلك المجموعات، ثم العمل على إنشاء تركيب خطي (أو غير خطي)  $F(x)$  لتلك المتحولات المستقلة، واستخدامه في التمييز بين تلك المجموعات التي يتألف من المجتمع المدروس .

ويمكن تمثيل التأثير المشترك للمتغيرات المستقلة  $X$  على المجموعات التابعة  $Y$  بيانياً على الشكل التالي:



الشكل (1-1) آلية عمل التحليل التمييزي

إن الشكل (1-1) يبين أن جملة المتغيرات المستقلة  $(X_1 X_2 \dots X_p)$  تؤثر على كل مجموعة من المجموعات التي يتألف منها التابع  $Y$ . وذلك من خلال التراكيب الخطية (أو غير الخطية)، المعرفة عليها  $F_1(x), F_2(x), \dots, F_g(x)$ ، وأن كل عنصر من عناصر هذه المجموعات يعطينا جملة من القياسات لهذه المتغيرات وعند تعويضها في التركيب الخطي المقابل لتلك المجموعة نحصل على قيمة محددة للتابع  $F_j(x)$  تميز ذلك العنصر عن غيره في المجموعة المذكورة .

إن تطبيق التحليل التمييزي يتطلب أن نقوم بسحب عينات عشوائية طبقية من كل مجموعة من مجموعات المجتمع  $Y$  بحجم متساوية أو مختلفة  $n_1, n_2, \dots, n_g$  فنتشكل لدينا عينة كلية حجمها  $n = \sum_{j=1}^g n_j$  . ويمكن تفريغ بيانات هذه العينات في جداول مناسبة وحساب متوسطاتها وتبايناتها كما يلي:

جدول (1-1) بيانات العينات المسحوبة من المجموعات (حالة ثلاث مجموعات):

البيانات المتحولات	بيانات عينة المجموعة الأولى من $n_1$ عنصراً 1 2 3 ... $n_1$	المتوسطات البيانات	التباين	بيانات عينة المجموعة الثانية من $n_2$ عنصراً 1 2 ... $n_2$	المتوسطات	التباين	بيانات عينة المجموعة الثالثة من $n_3$ عنصراً فيها للقيم $x_{ijk}$
$X_1$	$x_{111}, x_{112}, x_{113} \dots x_{11n_1}$	$\bar{x}_{11}$	$\sigma_{11}^2$	$x_{121}, x_{122} \dots x_{12n_2}$	$\bar{x}_{12}$	$\sigma_{12}^2$	نرمز لقيم المتحولات بـ $x_{ijk}$ حيث يرمز $i$ للمتحولات $X_i$ ويرمز $j$ للمجموعة $G_j$ ويرمز $k$ للملاحظة $k$ من $n_j$ وحيث أن: $i: 1 2 3 \dots p$ $j: 1 2 3 \dots g$ $k: 1 2 3 \dots n_j$
$X_2$	$x_{211}, x_{212}, x_{213} \dots x_{21n_1}$	$\bar{x}_{21}$	$\sigma_{21}^2$	$x_{221}, x_{222} \dots x_{22n_2}$	$\bar{x}_{22}$	$\sigma_{22}^2$	
$X_3$	$x_{311}, x_{312}, x_{313} \dots x_{31n_1}$	$\bar{x}_{31}$	$\sigma_{31}^2$	$x_{321}, x_{322} \dots x_{32n_2}$	$\bar{x}_{32}$	$\sigma_{32}^2$	
$X_4$	$x_{411}, x_{412}, x_{413} \dots x_{41n_1}$	$\bar{x}_{41}$	$\sigma_{41}^2$	$x_{421}, x_{422} \dots x_{42n_2}$	$\bar{x}_{42}$	$\sigma_{42}^2$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$X_p$	$x_{p11}, x_{p12}, x_{p13} \dots x_{p1n_1}$	$\bar{x}_{p1}$	$\sigma_{p1}^2$	$x_{p21}, x_{p22} \dots x_{p2n_2}$	$\bar{x}_{p2}$	$\sigma_{p2}^2$	

وأخيراً نشير إلى أننا سنركز في منشورنا هذا على التحليل التمييزي الخطي المتعدد الذي يعتمد على المتحولات المستقلة الكمية، ونعرفه كما يلي:

**التحليل التمييزي الخطي:** هو الذي يستخدم نماذج خطية في تركيب المتحولات المستقلة الكمية  $X_1 X_2 \dots X_p$  لاستخدامها في التمييز بين عناصر المجموعات . وهذه النماذج يكون لها الشكل التالي :

$$F(x) = B_0 + B_1X_1 + B_2X_2 + \dots + B_pX_p + \varepsilon \quad (1 - 1)$$

حيث أن  $B_0 B_1 B_2 \dots B_p$  هي الأمثال التمييزية ( المعاملات ) في التابع  $F(x)$  وأن  $\varepsilon$  هو متحول الخطأ العشوائي في النموذج (1-1) . وأن  $F(x)$  هو تابع كامن يصاغ بواسطة تركيب خطي للمتحولات  $X$  المؤثرة في التابع  $Y$  ويسمى بالتابع التمييزي الفاصل بين المجموعات (Discriminant Function) .

إن الهدف الأساسي للتحليل التمييزي هو اختبار فيما إذا كان تصنيف عناصر المجموعات في التابع  $Y$ ، يعتمد على متحول واحد على الأقل من المتحولات المستقلة  $X_1 X_2 \dots X_p$ ، ويمكن وضع فرضية العدم والفرضية البديلة كما يلي:

$H_0$ : إن تصنيف عناصر المجموعات في التابع  $Y$  لا يعتمد على أي من المتحولات المستقلة  $X_1 X_2 X_3 \dots X_p$  .

$H_1$ : إن تصنيف عناصر المجموعات في التابع  $Y$  يعتمد على واحد على الأقل من المتحولات المستقلة  $X_1 X_2 \dots X_p$  .

ويمكن كتابة هاتين الفرضيتين كما يلي :

$$\begin{aligned} H_0: & B_i = 0 & i : 1 2 3 \dots p & \text{من أجل} \\ H_1: & B_i \neq 0 & \text{من أجل قيمة واحدة عن الأقل} \end{aligned} \quad (2 - 1)$$

### 3-1 الافتراضات والشروط لتطبيق التحليل التمييزي:

يشترط عند تطبيق التحليل التمييزي على مجموعات مجتمع ما باستخدام متحولات  $X_1 X_2 X_3 \dots X_p$  مايلي :

1- أن تكون المجموعات في المجتمع المدروس محددة ومتكاملة وغير متقاطعة فيما بينها، وأن يكون عددها  $(g \geq 2)$ ، وأن تكون حجوماً متقاربة أو غير مختلفة كثيراً . ويمكن إنشاء المجموعات على أسس ديموغرافية أو جغرافية أو إدارية أو غيرها .

2- أن تكون المشاهدات (أو القياسات) مستقلة عن بعضها البعض . أي ان تكون العينات التطبيقية مسحوبة بطريقة عشوائية من المجموعات . وأن يكون حجم العينة الكلية  $n$  أكبر بعشرين مرة من عدد المتحولات المستقلة  $p$  [Steven 1996]. وأن لا يقل حجمها في كل الأحوال عن  $n = 30$  عنصراً .

3- أن تكون المتحولات  $X_1 X_2 X_3 \dots X_p$  مستقلة عن بعضها البعض . وأن لا يكون عددها كبيراً، وإذا كان عددها كبيراً فيجب اختزالها حسب معايير محددة . وفي كل الأحوال يجب أن لا يزيد عددها  $p$  عن 5% من حجم العينة الكلية  $n$ ، وعندما يكون  $p = 1$  فإننا نكون أمام التحليل التمييزي البسيط .

4- أن تكون المتحولات المستقلة (الكمية)  $X_1 X_2 \dots X_p$  خاضعة للتوزيع الطبيعي المتعدد، أو أن يكون كل منها خاضعاً للتوزيع الطبيعي العام . وإذا كان هذا الشرط غير محقق نقوم بإجراء بعض التحويلات اللازمة أو نستغني عن بعض المتحولات، ولكن إذا كان حجم العينة كبيراً فيطبق عليها قانون الأعداد الكبيرة وتعتبر طبيعية تقاربياً .

5- أن تكون مصفوفات التباين والتباين المشترك للمتحولات المستقلة  $X$  داخل كل مجموعة، متشابهة (متماثلة Similarity)، أي أن تكون تلك المصفوفات متساوية أو متجانسة على الأقل .

6- أن تكون مقادير الأخطاء (البواقي) موزعة عشوائياً. أي أن يكون توقعها مساوياً للصفر . ملاحظة: يمكن التساهل في هذه الشروط عندما يكون حجم العينة كبيراً .

وللتأكد من تحقق هذه الشروط يجب إجراء الاختبارات اللازمة لذلك . وسنعرضها لاحقاً في فقرة خاصة .

### 1-4 أهداف وخطوات التحليل التمييزي :

يهدف التحليل التمييزي إلى تحقيق عدة أغراض أهمها:

- 1- التحري عن الاختلافات الموجودة بين المجموعات التي يتألف منها المجتمع المدروس .
- 2- تحديد الطريقة الأكثر توفيراً في التفريق أو الفصل أو التمييز بين المجموعات .
- 3- حذف المتحولات المرتبطة بضعف مع المجموعة المحددة .
- 4- تحديد المتحولات المستقلة التي تميز كل مجموعة وأي المتحولات التي لا تميزها .
- 5- تصنيف العناصر أو الحالات ضمن المجموعات حسب قاعدة معينة .
- 6- اختبار النظرية القائمة من خلال مقارنة فيما إذا كانت الحالات المصنفة سابقاً مشابهة للحالات المصنفة لاحقاً .

أما خطوات التحليل التمييزي فتتلخص بما يلي :

- 1- تحديد نوع التحليل التمييزي المناسب الذي يوصلنا إلى النتائج الاحصائية المطلوبة، فهل التحليل الوصفي هو المناسب أم التنبؤي، أو البسيط أم المتعدد، أو الخطي أم غير الخطي، أو الكمي أم النوعي أم اللوجستي .
- 2- تحديد المتحولات المستقلة المناسبة للتحليل التمييزي المختار واللازمة لتحقيق أهداف البحث وجمع البيانات عنها .
- 3- تحديد عدد المجموعات في المجتمع التي ستستخدم في التحليل ( $g \geq 2$ ) وسحب العينات التطبيقية منها وجمع البيانات اللازمة منها .
- 4- اختبار البيانات والتأكد من أنها تحقق الافتراضات والشروط المفروضة عليها .
- 5- إجراء التحليل التمييزي حسب خطواته العملية والحصول على النتائج المطلوبة .
- 6- تفسير النتائج والعمل على الاستفادة منها .