الفصل الثالث

التحليل التمييزي الخطي (بمتحولين ومجموعتين)

3-1 : تمهيد:

لنأخذ المثال الوارد في التمهيد (2-1) والذي يتضمن القروض المصرفية الممنوحة لزبائن أحد المصارف. ولنفترض الآن أن هذه القروض مصنفة إلى مجموعتين متعثرة G_1 وغير المتعثرة G_2، ولنفترض إننا نريد أن نميز بينهما باستخدام متحولين هما:

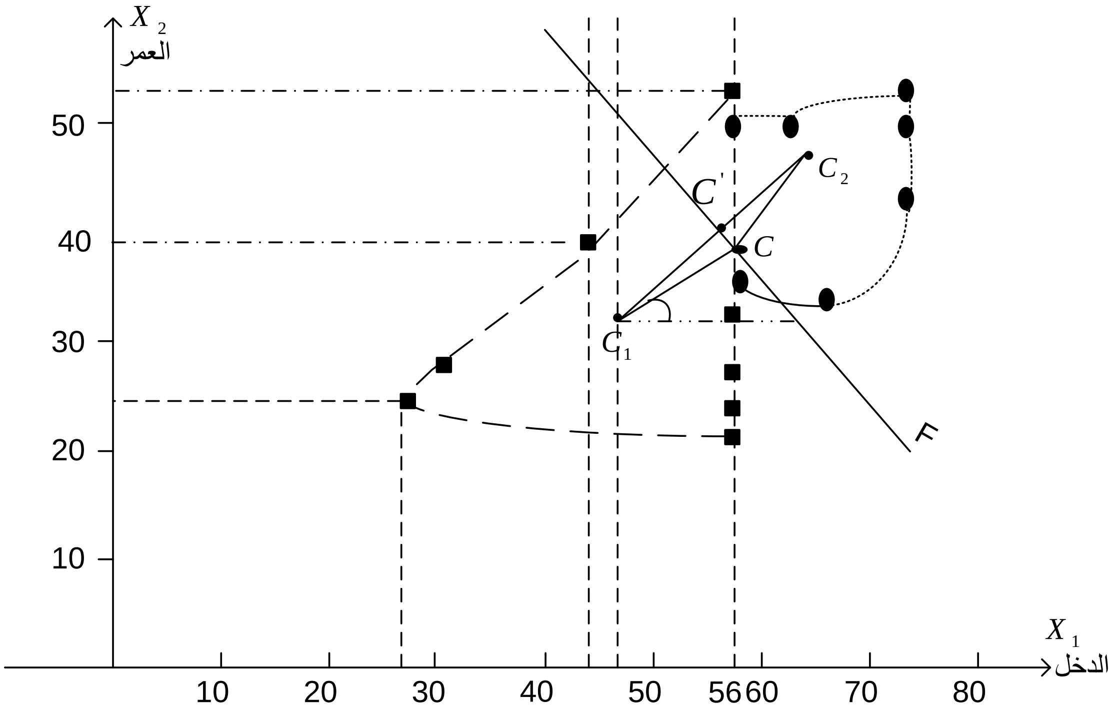
- الدخل الشهري لصاحب القرض ونرمز له بـ X_1 (بالآلاف) .
- العمر الذي يبلغه صاحب القرض عند المنح ونرمز له بـ X_2 (بالسنوات)

ولنفترض أننا سحبنا عينتين عشوائيتين من هاتين المجموعتين بحجمين $n_1 = 8$ و $n_2 = 7$ فحصلنا منهما على البيانات المبينة في الجدول التالي:

جدول (3-1): بيانات عينتي القروض . (فرضية)

مجموعة القروض المعثرة G_1										مجموعة القروض غير المعثرة G_2									
عناصر عينة القروض المتعثرة $n_1 = 8$									المتوسط	عناصر عينة القروض غير المتعثرة $n_2 = 7$								المتوسط	
الرقم	1	2	3	4	5	6	7	8		الرقم	1	2	3	4	5	6	7		
الدخل X_1	56	56	45	56	29	56	56	30	48	الدخل X_1	72	72	56	72	64	64	56	65.14	
العمر X_2	52	23	40	34	26	25	27	29	32	العمر X_2	54	50	48	44	47	35	36	44.86	
إشارة F	+	−	−	−	−	−	−	−	−	إشارة F	+	+	+	+	+	+	−	+	
D_1^2	4.6	1.29	6.88	....	...	....	...	2.33		D_1^2	8.95	...	...	...	...	....	0.51		
D_2^2	2.6	10.14	8.30	...	...	...	...	28.84		D_2^2	1.57	...	...	...	...	...	31.8		
التصنيف حسب المسافة الاصغر	2	1	1	1	1	1	1	1	1	التصنيف حسب المسافة الاصغر	2	2	2	2	2	2	1	2	

وعند رسم هذه القروض كنقاط هندسية إحداثياتها (X_1 , X_2) في المستوي جعلنا المحور الأفقي للدخل X_1 والعمودي للعمر X_2 فحصلنا على الشكل البياني التالي:



الشكل (3−1) مواقع عناصر المجموعتين: ■عناصر G_1 ، ● عناصر G_2 ومراكزهما

ومن البيانات السابقة نرسم مركزي هاتين المجموعتين:

C_1 مركز المجموعة الأولى G_1 وإحداثياته $C_1(48\,,32)$.

C_2 مركز المجموعة الثانية G_2 وإحداثياته $C_2(65.14\,,44.86)$.

كما يمكننا حساب ورسم المركز العام لهذه القروض. ونرمز له بـ C ونحسب إحداثياته من متوسطي الدخل والعمر في العينتين معاً فنحصل على أن $C(56\,,38)$، كما يمكننا حساب القيمة الوسطى لمركزي المجموعتين من متوسطي إحداثيات المركزين السابقين فنحصل على القيمة الوسطى للمركزين ونرمز له بـ C' وتكون إحداثياته :

$$C'\left(\frac{48+65.14}{2}\,,\frac{32+44.86}{2}\right)=C'(56.57\,,38.43)$$

ملاحظة: إن القيمة الوسطى C' تقع على القطعة المستقيمة $C_1\,,C_2$ وتقع في منتصفها. ولكن المركز العام C يقع خارج القطعة المستقيمة $C_1\,,C_2$ كما موضح على الشكل (3−1) .

والسؤال الآن كيف يمكننا أن نصيغ علاقة رياضية للتمييز بين هاتين المجموعتين؟

الجواب يمكن أن يكون بعدة طرائق رياضية هي:

3−2 : طريقة المستقيم الفاصل بين عناصر المجموعتين، ويسمى تابع التمييز القانوني Canonical Discriminant Function

تتلخص هذه الطريقة بإنشاء مستقيم يمر من نقطة القيمة الوسطى للمركزين $C'(56.57\,,38.43)$ (أو من أي نقطة مناسبة أخرى)، ويتعامد (أو يتقاطع) مع القطعة المستقيمة الواصلة بين مركزي المجموعتين

$C_1(48\,,32)$ و $C_2(65.14\,,44.86)$، فنحصل على مستقيم يفرز عناصر العينة إلى مجموعتين جديدتين منفصلتين: المجموعة الأولى تقع تحته، والمجموعة الثانية تقع فوقه .

ولإيجاد معادلة هذا المستقيم الفاصل بين هاتين المجموعتين نقوم بإيجاد ميل القطعة المستقيمة $\overline{C_1\,C_2}$ الواصلة بين مركزي المجموعتين من العلاقة :

$$m = tg\theta = \frac{44.86 - 32}{65.14 - 48} = \frac{12.86}{17.14} = 0.75029$$

ثم نقوم بحساب ميل المستقيم المتعامد معها من العلاقة:

$$m' = -\frac{1}{m} = \frac{-1}{0.75020} = -1.3328$$

ثم نقوم بحساب معادلة المستقيم المعامد، والذي يمر من نقطة القيمة الوسطى للمركزين $C'(56.57\,,38.43)$ من العلاقة :

$$(X_2 - 38.43) = -1.3328(X_1 - 56.57)$$

وبعد الإصلاح نحصل على معادلة المستقيم التالي:

$$X_2 + 1.3328X_1 - 113.8265 = 0 \qquad (1-3)$$

وهنا نلاحظ أن هذا المستقيم يقسم نقاط المستوى إلى نصفين: نصف موجب يقع فوقه ونصف سالب يقع تحته وجزء معدوم يقع عليه . كما أن هذا المستقيم يفرز عناصر العينة إلى مجموعتين منفصلتين:

المجموعة الأولى: هي النقاط التي تقع تحته (في النصف السالب) وتضم معظم نقاط القروض المتعثرة في G_1 [ماعدا القرض الأول (52 , 56)] .

المجموعة الثانية: وهي النقاط التي تقع فوقه (في النصف الموجب) ويضم معظم نقاط القروض غير المتعثرة في G_2 (ماعدا القرض الأخير (36 , 56) .

أي أنه يمكننا استخدام هذا المستقيم (أو أي مستقيم أفضل منه) كأداة لفرز القروض إلى متعثرة وغير متعثرة باحتمال خطأ معقول .

ولذلك نفترض أن معادلة المستقيم تساوي تابعاً F ونكتبها كمايلي:

$$F = X_2 + 1.3328X_1 - 113.8265 \qquad (2-3)$$

ولاستخدام هذه المعادلة في تصنيف عناصر العينة أو في تصنيف أي عنصر جديد من المجتمع. نقوم بتعويض إحداثيات صاحب القرض$(X_1\,,X_2)$ في المعادلة $(3-2)$ ونحسب قيمة F ثم نتخذ القرار (في هذا المثال) حسب إشارة قيمة F كما يلي:

إذا كانت $F \leq 0$ فإن القرض ينتمي إلى المجموعة المتعثرة G_1 .

أما إذا كانت $F > 0$ فإن القرض ينتمي إلى المجموعة غير المتعثرة G_2 .

وإذا طبقنا ذلك على عناصر العينتين المسحوبتين من $(G_1\,,G_2)$، نحصل على السطر F في الجدول $(3-1)$. وإذا قارنا التصنيف الجديد مع التصنيف الفعلي أو الإداري السابق نحصل على الجدول التالي:

جدول (3-2): نتائج التصنيف الجديد مع التصنيف الأصلي:

	التصنيف الجديد			
التصنيف الفعلي الإداري	المجموعات	G_1	G_2	المجموع
	G_1	7	1	8
	G_2	1	6	7

وهنا نلاحظ أن نسبة التصنيف الصحيح تساوي نسبة مجموع عدد العناصر القطرية على حجم العينة الكلية n، أي:

$$P = \frac{7+6}{15} = 0.8667 = (86.67\ \%)$$

وإن نسبة التصنيف الخاطئ تساوي نسبة مجموع العناصر غير القطرية على حجم العينة الكلية n،، أي:

$$q = \frac{1+1}{15} = 0.1333 = (13.33\ \%)$$

وهذا يعني أنه يمكننا الاعتماد على المستقيم السابق لتصنيف عناصر المجتمع مع خطأ معقول .

ملاحظة: إن هذا المستقيم الفاصل بين هاتين المجموعتين ليس هو الفاصل المثالي بينهما، بل هو أحد المستقيمات الفاصلة بينهما، لأنه يمكن أن ننشأ مستقيمات كثيرة أخرى تفصل بين هاتين المجموعتين وبطريقة أفضل. فمثلاً يمكننا أن نحرك هذا المستقيم يميناً أو يساراً، أو إلى الأعلى أو إلى الأسفل، لنحصل على لانهاية من المستقيمات الفاصلة بين هاتين المجموعتين. كما يمكننا أن نغير ميلان هذا المستقيم m' إلى الأكبر أو إلى الأصغر، وندوره حول نقطة القيمة الوسطى نحو اليمين أو نحو اليسار بزاوية θ، لنحصل على لانهاية أخرى من المستقيمات الفاصلة بين هاتين المجموعتين .

لذلك كان لابد لنا من أن نضع أسس رياضية دقيقة لإنشاء ذلك المستقيم الفاصل بين هاتين المجموعتين بحيث نجعل الاحتمال الاجمالي للتصنيف الخاطئ TPM أصغر ما يمكن، أو نجعل القيمة المتوقعة لتكاليف التصنيف الخاطئ ECM أصغر ما يمكن. وهذا ما سنتعرض إليه لاحقاً .

وأخيراً نشير إلى أنه عندما تكون المجموعات متقاطعة ومتشابكة، فإنها تكون غير قابلة للفصل بواسطة خط مستقيم، وتسمى مثل هذه المجموعات بالمجموعات غير القابلة للفصل خطياً، وفي هذه الحالة نلجأ إلى استخدام منحنيات مناسبة للفصل بينها، وبذلك ينتقل التحليل إلى التحليل التمييزي غير الخطي، وهو موضوع خارج إطار هذا المنشور .

3-3 : طريقة الجوار الأقرب (أصغر المسافات عن مراكز المجموعات):

تعتمد هذه الطريقة على حساب مربع المسافة بين كل عنصر x_0 ومركزي المجموعتين $(G_1\ , G_2)$، واللذين يساويان حسب الجدول (3-1) ما يلي:

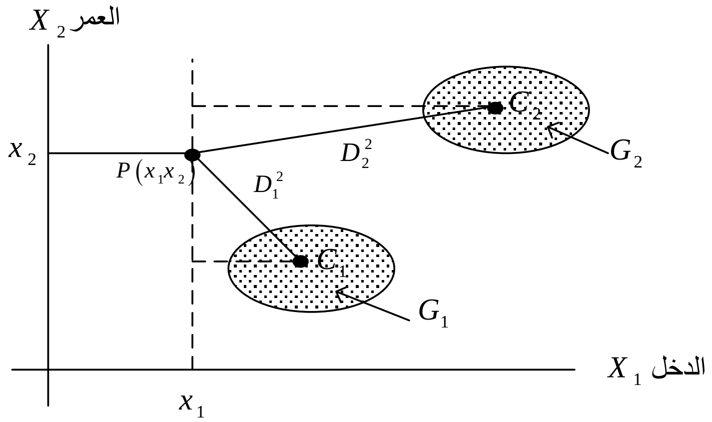
جدول (3-3): إحداثيات مركزي المجموعتين بالرموز والأعداد:

	G_1	G_2
الدخل X_1	48	65.14
العمر X_2	32	44.86

=

	G_1	G_2
X_1	$\bar{x}_{11}$	$\bar{x}_{12}$
X_2	$\bar{x}_{21}$	$\bar{x}_{22}$

لنأخذ الآن نقطة ممثلة لأحد القروض الجديدة مثل النقطة P والتي إحداثياتها يساويان $P(x_1 , x_2)$، ولنرمز لمربعي المسافة الهندسية منها حتى مركزي المجموعتين C_1 و C_2 بالرمزين D_1^2 و D_2^2 على الترتيب ونرسمها على الشكل البياني كما يلي:



الشكل (3-2): المسافات عن مركزي المجموعتين

ولتحديد المجموعة التي ينتمي إليها ذلك القرض الجديد نحسب مربعي المسافتين D_1^2 و D_2^2 ثم نقارن بينهما ونتخذ القرار حول الانتماء كما يلي:

إذا كانت $D_1^2 \leq D_2^2$ فإننا ننسب النقطة P إلى المجموعة G_1 .

أما إذا كانت $D_1^2 > D_2^2$ فإننا ننسب النقطة P إلى المجموعة G_2 .

ولكن هذه المقارنة لا تجوز إلا إذا كانت القيم العددية لهاتين المسافتين مقاسة بوحدات قياس موحدة، وبما أن X_1 – في مثالنا– يقاس بآلاف الليرات و X_2 يقاس بالسنوات، فلا يجوز مقارنة هذين المتحولين بطريقة مباشرة .

وللتخلص من هذه المشكلة نقوم بتحويل هذين المتحولين إلى متحولين معياريين، وذلك بإجراء التحويل المعياري $Z = \frac{X-\bar{X}}{\sigma}$، على كل متحول وفي كل مجموعة على حدة، فنحصل على قيم مجردة من الوحدات . ولتوضيح ذلك نحسب الانحرافات المعيارية لكل متحول وفي كل مجموعة على حدة، فنحصل على الانحرافات المعيارية التالية:

جدول (3-4): الانحرافات المعيارية لعناصر العينتين:

المجموعات / المتحولات	G_1	G_2
X_1	12.036	7.198
X_2	9.769	7.081

=

المجموعات / المتحولات	G_1	G_2
X_1	S_{11}	S_{12}
X_2	S_{21}	S_{22}

وبذلك نجد أن مربع المسافة المعيارية بين النقطة$P(X_1\,,X_2)$ ومركز المجموعة الأولى $C_1(48.32)$ يحسب من العلاقة:

$$D_1^2=\left(\frac{X_1-48}{12.036}\right)^2+\left(\frac{X_2-32}{9.769}\right)^2 \qquad (3-3)$$

وإن مربع المسافة المعيارية بين النقطة$P(X_1\,,X_2)$ ومركز المجموعة الثانية $C_2(65.14\,,44.86)$ يحسب من العلاقة:

$$D_2^2=\left(\frac{X_1-65.14}{7.198}\right)^2+\left(\frac{X_2-44.86}{7.081}\right)^2 \qquad (4-3)$$

ومنهما نحصل على قيم D_1^2 و D_2^2 كقيم عددية مجردة من وحدات القياس . لذلك يمكن مقارنتهما وتحديد المجموعة التي تنتمي إليها النقطة P ، وهي المجموعة التي تقابل أصغر المربعين D_1^2 أو D_2^2 .

فمثلاً نجد أن مربعي المسافة للقرض الأول من G_1 وهو الذي يقابل النقطة $P(56\,,52)$ يساويان :

$$D_1^2=\left(\frac{56-48}{12.036}\right)^2+\left(\frac{52-32}{9.769}\right)^2=0.44418+4.1914=4.6332$$

$$D_2^2=\left(\frac{56-65.14}{7.198}\right)^2+\left(\frac{52-44.86}{7.081}\right)^2=1.6124+1.0167=2.6291$$

وعند المقارنة نجد أن:$D_2^2<D_1^2$ وهذا يعني أن النقطة Pهي أقرب إلى مركز المجموعةG_2 لذلك سنعتبر ذلك القرض قرضاً غير متعثر وينتمي إلى المجموعة G_2، مع أنه ينتمي إلى G_1(**انظر الجدول (3−1)**).

وهكذا قمنا بحساب مربعات المسافات المقابلة لبعض عناصر العينة (عينة القروض) في المثال السابق ووضعناها في سطرين ملحقين بالجدول (3−1)، وبناء على مقارنة D_1^2 و D_2^2 في هذين السطرين تم تحديد انتماء كل عنصر من عناصر عينة البحث، ووضعناها في السطر الأخير، فحصلنا على نفس النتائج التي حصلنا عليها بطريقة المستقيم الفاصل المبينة في الجدول (3−2).

ويمكن الاستفادة من هذه الطريقة في تحديد انتماء أي قرض جديد والتنبؤ بمصيره قبل اتخاذ قرار الموافقة على منحه .

ملاحظة: لنأخذ المجموعتين الجديدتين G_1' وG_2'، اللتين أصبحتا بعد التصنيف الجديد يضمان العناصر التالية:

$$G_1':\{7_{(1)},2,3,4,5,6,7,8\}$$

$$G_2':\{1,2,3,4,5,6,\ 8_{(1)}\}$$

وإذا حسبنا متوسطات X_1 وX_2 في هاتين المجموعتين الجديدتين نحصل على المتوسطات والتباينات التالية:

المتوسطات الجديدة للمجموعات				الانحرافات المعيارية الجديدة		
	G_1	G_2			G_1	G_2
X_1	48	56		X_1	12.036	7.198
X_2	30	47.14		X_2	6	6.283

أي أن إحداثيات مركز المجموعة الجديدة G_1' أصبح $C_1'(48,30)$.

وأن إحداثيات مركز المجموعة الجديدة G_2' أصبح $C_2'(56,47.14)$.

وإن مربع المسافة الهندسية بينهما يساوي:

$$D_1'^2 = (48-56)^2 + (30-47.11)^2 = 357.78$$

بينما كان مربع المسافة بينهما حسب التصنيف الفعلي (الإداري) يساوي مربع البعدين المركزين الفعليين اللذين إحداثياتها $C_1(48,32)$ $C_2(56,44.86)$ ويساوي:

$$D_1^2 = (48-56)^2 + (32-44.86)^2 = 211.38$$

وهكذا نلاحظ أن التصنيف الجديد أدى إلى جعل مربع المسافة بين مركزي المجموعتين أكبر مما كان عليه قبل ذلك. وأصبحت (357.78) بدلاً من (211.38) . ولقد استفاد (فيشر) من هذه الملاحظة واستنبط معياراً خاصاً للتمييز بين عناصر المجموعتين G_1 و G_2.

كما نلاحظ أن التصنيف السابق جعل مجموع مربعات الانحرافات داخل كل مجموعة SSW يصبح أصغر أو يساوي مما كان عليه، وبالمقابل جعل مجموع مربعات الانحرافات بين هاتين المجموعتين SSB يصبح أكبر مما كان عليه .

وهذا يعني أن المجموعتين أصبحتا أكثر تجانساً داخلياً وأكثر اختلافاً خارجياً، وهو ما نسعى إليه من خلال التحلل التمييزي بين المجموعات .

3-4 طريقة القيمة المتوقعة للتكاليف ECM (لمتحولين طبيعيين (X_1, X_2) ومجموعتين $((G_1, G_2)$: [Johnson A. R. P.485 بتصرف] .

قبل البحث في التحليل التمييزي لهذين المتحولين نرمز لهما بـ X ولتوقعهما بـ $\boldsymbol{\mu}$ ونكتبهما على شكل شعاعين عمودين كما يلي:

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad (5-3)$$

وبذلك نجد أن التوقع الرياضي لهما في المجتمع ككل يساوي:

$$E(X) = \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \boldsymbol{\mu} \qquad (6-3)$$

وعندها فإن مصفوفة التباين المشترك لهما تساوي:

$$V = cov\,(X_1, X_2) = E(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})'$$
$$= E\begin{bmatrix} X_1-\mu_1 \\ X_2-\mu_2 \end{bmatrix} * [(X_1-\mu_1), (X_2-\mu_2)]$$
$$V = E\begin{bmatrix} (X_1-\mu_1)^2 & (X_1-\mu_1)(X_2-\mu_2) \\ (X_2-\mu_2)(X_1-\mu_1) & (X_2-\mu_2)^2 \end{bmatrix} = \begin{bmatrix} E(X_1-\mu_1)^2 & E(X_1-\mu_1)(X_2-\mu_2) \\ E(X_2-\mu_2)(X_1-\mu_1) & E(X_2-\mu_2)^2 \end{bmatrix}$$
$$V = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \qquad (7-3)$$

حيث أن: σ_{11} هو تباين X_1 و σ_{22} هو تباين X_2

وأن: $\sigma_{12} = \sigma_{21}$ هو التباين المشترك للمتحولين (X_1, X_2). وأن V هي مصفوفة مربعة متناظرة .

ومن العلاقة $(7-3)$ يمكننا أيضاً حساب المصفوفة الارتباطية المتناظرة للمتحولين (X_1 , X_2) والتي تعتمد على أن معامل الارتباط بينهما ρ_{12} يحسب من العلاقة: $\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}} * \sqrt{\sigma_{22}}}$ وبذلك يكون لدينا:

$$\rho = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}} * \sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}} * \sqrt{\sigma_{22}}} \\ \frac{\sigma_{21}}{\sqrt{\sigma_{22}} * \sqrt{\sigma_{11}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}} * \sqrt{\sigma_{22}}} \end{bmatrix} = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{bmatrix} \qquad (8-3)$$

حيث أن: $\rho_{12} = \rho_{21}$ هو معامل الارتباط بين المتحولين (X_1 , X_2)

ومن المصفوفة $(7-3)$ يمكننا استخلاص مصفوفة الانحرافات المعيارية الخاصة بهذين المتحولين، والتي سنرمز لها (تجاوزاً) بالرمز $\sigma^{\frac{1}{2}}$. وهي المصفوفة القطرية المؤلفة من جذور العناصر القطرية في المصفوفة V ونكتبها كما يلي:

$$\sigma^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 \\ 0 & \sqrt{\sigma_{22}} \end{bmatrix} \qquad (9-3)$$

وبسهولة يمكننا أن نستخلص أن V تساوي:

$$V = \sigma^{\frac{1}{2}} * \rho * \sigma^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 \\ 0 & \sqrt{\sigma_{22}} \end{bmatrix} * \begin{bmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{bmatrix} \begin{bmatrix} \sqrt{\sigma_{11}} & 0 \\ 0 & \sqrt{\sigma_{22}} \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{\sigma_{11}} & 0 \\ 0 & \sqrt{\sigma_{22}} \end{bmatrix} * \begin{bmatrix} \sqrt{\sigma_{11}} & \rho_{12}\sqrt{\sigma_{22}} \\ \rho_{21}\sqrt{\sigma_{11}} & \sqrt{\sigma_{22}} \end{bmatrix} = \begin{bmatrix} \sigma_{11} \ , & \rho_{12}\sqrt{\sigma_{11}} * \sqrt{\sigma_{22}} \\ \rho_{21}\sqrt{\sigma_{22}} * \sqrt{\sigma_{11}} & , \ \sigma_{22} \end{bmatrix}$$

وبتبديل ρ_{12} و ρ_{21} من العلاقة: $\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}$ نحصل على أن:

$$V = \sigma^{\frac{1}{2}} * \rho * \sigma^{\frac{1}{2}} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \qquad (10-3)$$

وإذا ضربنا طرفي العلاقة $(10-3)$ من اليمين ثم من اليسار بـ $\sigma^{-\frac{1}{2}}$ نحصل على أن:

$$\rho = \sigma^{-\frac{1}{2}} * V * \sigma^{-\frac{1}{2}} \qquad (11-3)$$

حيث أن: $\sigma^{-\frac{1}{2}}$ هي مقلوب مصفوفة الانحرافات المعيارية $\sigma^{\frac{1}{2}}$، وهي مصفوفة قطرية تكون عناصرها مساوية لمقاليب عناصر المصفوفة $\sigma^{\frac{1}{2}}$ أي أن:

$$\sigma^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} \end{bmatrix} \qquad \mathbf{(12-3)}$$

ويمكن تعميم ذلك على عدة متحولات $X_1, X_2, X_3, \dots, X_p$ وعلى عدة مجموعات $G_1, G_2, \dots, G_p$ كما سنرى لاحقاً.

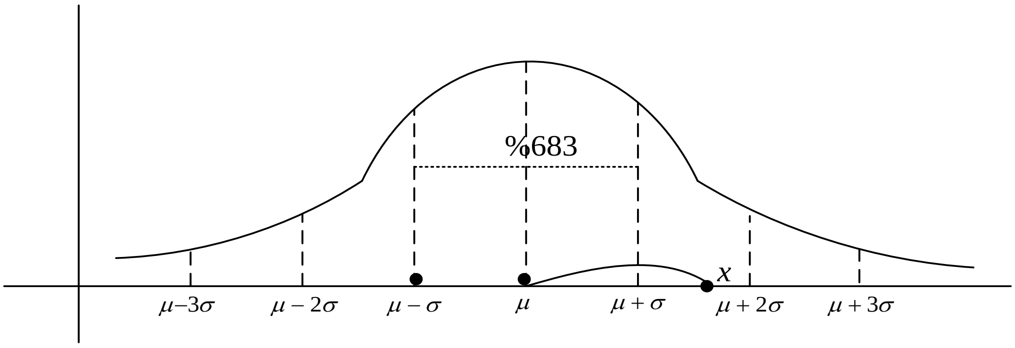
ولتطبيق طريقة القيمة المتوقعة للتكاليف على هذه الحالة (لمتحولين طبيعيين X_1 و X_2 وعلى المجموعتين G_1 و G_2) نفترض أن التوزيع الطبيعي لـ X_1 هو $N(\mu_1 \ \ \sigma_1^2)$. والذي يعطي بالعلاقة :

$$f(x_1) = \frac{1}{\sqrt{2\pi} * \sigma_1} * e^{-\frac{1}{2}\left(\frac{X_1 - \mu_1}{\sigma_1}\right)^2} : \ -\infty < X_1 < +\infty \qquad (13-3)$$

وإن التوزيع الطبيعي لـ X_2 هو $N(\mu_2 \ \sigma_2^2)$ وهو يعطي بالعلاقة :

$$f(x_2) = \frac{1}{\sqrt{2\pi} * \sigma_2} * e^{-\frac{1}{2}\left(\frac{X_2 - \mu_2}{\sigma_2}\right)^2} : \ -\infty < X_2 < +\infty \qquad (14-3)$$

وإن كل منها يرسم الشكل البياني التالي :



الشكل (3-3): المنحني العام للتوزيع الطبيعي

وهنا نلاحظ أنه يمكن كتابة الحد الأسي في كل منهما على الشكل التالي :

$$D^2 = \left(\frac{X - \mu}{\sigma}\right)^2 = (X - \mu) * (\sigma^2)^{-1} * (X - \mu) \qquad (15-3)$$

وهو يمثل مربع المسافة الإحصائية بين النقطة X حتى المتوسط μ مقاسة بواحدات الانحراف المعياريσ.

ويمكننا تعميم هذه العلاقة على أي شعاع X مؤلف من P متحولاً وذات متوسط عام μ ومصفوفة تباين مشترك V كما يلي:

$$\Delta^2 = (\boldsymbol{X} - \boldsymbol{\mu})' * \ \boldsymbol{V}^{-1} * (\boldsymbol{X} - \boldsymbol{\mu}) \qquad (16-3)$$

حيثΔ : هو مربع المسافة الإحصائية من X حتى μ في الفضاء R^p .

وبناء على ذلك يمكننا تعريف التوزيع المشترك للمتحولات $X_1, X_2, \dots, X_p$ باستبدال مربع المسافة D^2 في (3-15) بمربع المسافة Δ^2 ، فنحصل على العلاقة التالية :

$$f(X_1, X_2, \dots, X_p) = \frac{1}{(2\pi)^{\frac{P}{2}} |V|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{X} - \boldsymbol{\mu})' * \boldsymbol{V}^{-1} * (\boldsymbol{X} - \boldsymbol{\mu})} \qquad (16\,a-3)$$

ولإيجاد التوزيع المشترك للمتحولين (X_1, X_2) نستبدل كل X بالشعاع $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ والتوقع μ بالشعاع $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ ، وكما نستبدل مقلوب مصفوفة التباين V^{-1} بما تساويه والذي نحسبه من V كما يلي:

بما أن V تساوي:

$$V = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \qquad |V| = \sigma_{11}\,\sigma_{22} - \sigma_{12}\,\sigma_{21} \qquad (17-3)$$

فإن مقلوبها V^{-1} يساوي:

$$V^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - {\sigma_{12}}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{21} & \sigma_{11} \end{bmatrix} \qquad (18-3)$$

ونكتب التوزيع المشترك للمتحولين (X_1, X_2) قياساً على العلاقة $(16\,a-3)$ كمايلي:

$$f(X) = f(X_1, X_2) = \frac{1}{(2\pi)|V|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{X} - \boldsymbol{\mu})' * \boldsymbol{V}^{-1} * (\boldsymbol{X} - \boldsymbol{\mu})} \qquad (19-3)$$

وإن هذا التوزيع المشترك يعرف على المجموعة الأولى G_1 بالعلاقة :

$$f_1(x)=\frac{1}{(2\pi)|V_1|^{\frac{1}{2}}}e^{-\frac{1}{2}}(\boldsymbol{X}-\boldsymbol{\mu_1})'*V_1{}^{-1}*(\boldsymbol{X}-\boldsymbol{\mu_1}) \qquad (20-3)$$

ويعرف على المجموعة الأولى G_2 بالعلاقة :

$$f_2(x)=\frac{1}{\sqrt{2\pi}|V_2|^{\frac{1}{2}}}e^{-\frac{1}{2}}(\boldsymbol{X}-\boldsymbol{\mu_2})'*V_2{}^{-1}*(\boldsymbol{X}-\boldsymbol{\mu_2}) \qquad (21-3)$$

وهنا سنميز بين حالتين هما:

- **الحالة الأولى:** عندما يكون لدينا $V_1=V_2=V$ افتراضياً أو فعلياً .

وهذا يعني إننا نفترض أن مصفوفتي التباين المشترك متساويتان $(V_1=V_2=V)$، وإن تكلفة التصنيف الصحيح معدومة $C(2/2)=C(1/1)=0$، وأن الاحتمال السابق للانتماء إلى المجموعة G_1 يساوي P_1، وإن الاحتمال السابق للانتماء إلى المجموعة G_2 يساوي P_2 .

وبذلك نجد أن قاعدة القيمة المتوقعة للتكاليف (2-21)، (2-22) تأخذ الشكل التالي:

$$R_1:\frac{f_1(x)}{f_2(x)}\geq\left[\frac{C(1/2)}{C(2/2)}\right]*\left[\frac{P_2}{P_1}\right] \qquad (22-3)$$

$$R_2:\frac{f_1(x)}{f_2(x)}<\left[\frac{C(1/2)}{C(2/1)}\right]*\left[\frac{P_2}{P_1}\right] \qquad (23-3)$$

ولذلك نحسب النسبة $\frac{f_1(x)}{f_2(x)}$ للتوزيعين الطبيعيين فنجد أنها عندما تكون $(V_1=V_2=V)$ تساوي:

$$\boldsymbol{\frac{f_1(x)}{f_2(x)}=\frac{\frac{1}{(2\pi)|V|^{\frac{1}{2}}}e^{-\frac{1}{2}}[(X-\mu_1)'*V^{-1}*(X-\mu_2)}{\frac{1}{(2\pi)|V|^{\frac{1}{2}}}e^{-\frac{1}{2}}[(X-\mu_2)'*V^{-1}*(X-\mu_2)}} \qquad \boldsymbol{(24-3)}$$

نختصر على العدد $\frac{1}{(2\pi)|V|^{\frac{1}{2}}}$ ونرفع المقام إلى البسط فنجد أن:

$$\boldsymbol{\frac{f_1(x)}{f_2(x)}=e^{-\frac{1}{2}}[(X-\mu_1)'*V^{-1}(X-\mu_1)]+\frac{1}{2}[(X-\mu_2)'*V^{-1}(X-\mu_2)]=e^{Z}} \quad \boldsymbol{(25-3)}$$

لنأخذ الأس العام ونرمز له بـ Z ونعالجه كما يلي:

$$\boldsymbol{Z=-\frac{1}{2}(X-\mu_1)'*V^{-1}(X-\mu_1)+\frac{1}{2}(X-\mu_2)'*V^{-1}(X-\mu_2)=}$$

$$\boldsymbol{=-\frac{1}{2}(X-\mu_1)'*V^{-1}X+\frac{1}{2}(X-\mu_1)'*V^{-1}\mu_1+\frac{1}{2}(X-\mu_2)'*V^{-1}X-\frac{1}{2}(X-\mu_2)'V^{-1}\mu_2=}$$

$$\boldsymbol{=-\frac{1}{2}[(X-\mu_1)'-(X-\mu_2)']V^{-1}X+\frac{1}{2}X'*V^{-1}\mu_1-\frac{1}{2}\mu_1'V^{-1}\mu_1-\frac{1}{2}X'*V^{-1}\mu_2+\frac{1}{2}\mu_2'V^{-1}\mu_2=}$$

$$\boldsymbol{=\frac{1}{2}(\mu_1-\mu_2)'*V^{-1}X+\frac{1}{2}[\mu_1'V^{-1}X]'-\frac{1}{2}[\mu_2'V^{-1}X]'-\frac{1}{2}\mu_1'V^{-1}\mu_1+\frac{1}{2}\mu_2'V^{-1}\mu_2=}$$

$$\boldsymbol{=\frac{1}{2}(\mu_1-\mu_2)'*V^{-1}X+\frac{1}{2}[(\mu_1-\mu_2)'*V^{-1}X]'-\frac{1}{2}[(\mu_1-\mu_2)'*V^{-1}(\mu_1+\mu_2)}$$

$$\boldsymbol{Z=(\mu_1-\mu_2)'*V^{-1}X-\frac{1}{2}(\mu_1-\mu_2)*V^{-1}(\mu_1+\mu_2)} \qquad \boldsymbol{(27-3)}$$

وذلك لأن الحد الثاني يساوي: عدد حقيقي $= [(\mu_1-\mu_2)' * V^{-1}X = [(\mu_1-\mu_2)' * V^{-1}X]'$ ويدمج مع الحد الأول .

وبذلك نجد أن القاعدة (3−22) للقيمة المتوقعة للتكاليف نأخذ الشكل التالي:

$$R_1: \frac{f_1(x)}{f_2(x)} = e^{(\mu_1-\mu_2)' * V^{-1}X - \frac{1}{2}(\mu_1-\mu_2)' * V^{-1}(\mu_1+\mu_2)} \geq \left[\left[\frac{C(1/2)}{C(2/1)}\right] * \left[\frac{P_2}{P_1}\right]\right] \quad \mathbf{(28-3)}$$

ثم نأخذ اللوغاريتم الطبيعي للطرفين فنجد أن القاعدة (3−22) تأخذ الشكل التالي:

$$R_1: (\boldsymbol{\mu_1}-\boldsymbol{\mu_2})' * V^{-1}X - \frac{1}{2}(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})' * V^{-1}(\boldsymbol{\mu_1}+\boldsymbol{\mu_2}) \geq \ell n\left[\left[\frac{C(1/2)}{C(2/1)}\right] * \left[\frac{P_2}{P_1}\right]\right] \quad \mathbf{(29-3)}$$

$$R_2: (\boldsymbol{\mu_1}-\boldsymbol{\mu_2})' * V^{-1}X - \frac{1}{2}(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})' * V^{-1}(\boldsymbol{\mu_1}+\boldsymbol{\mu_2}) < \ell n\left[\left[\frac{C(1/2)}{C(2/1)}\right] * \left[\frac{P_2}{P_1}\right]\right] \quad \mathbf{(30-3)}$$

وإذا لم تكن المعالم μ_1 وμ_2 وV معلومة نقوم بتقديرها من العينتين المسحوبتين من المجموعتين G_1 وG_2 بواسطة المؤشرات المقابلة لها $\bar{X}_1$ و$\bar{X}_2$ وبمصفوفة التباين المدمجة *Pooled* والمركبة من المصفوفتين S_1 وS_2 كما يلي:

$$S_p = \frac{(n_1-1)S_1 + (n_2-1)S_2}{n_1+n_2-2} \quad \mathbf{(31-3)}$$

وبالتالي نحصل على قاعدة القيمة المتوقعة للتكاليف في العينات من العلاقة التقديرية التالية:

$$R_1: (\bar{X}_1-\bar{X}_2)' * S_p^{-1} * X - \frac{1}{2}(\bar{X}_1-\bar{X}_2)' * S_p^{-1}(\bar{X}_1+\bar{X}_2) \geq \ell n\left[\frac{C(1/2)}{C(2/1)} * \frac{P_2}{P_1}\right] \quad \mathbf{(32-3)}$$

$$R_2: (\bar{X}_1-\bar{X}_2)' * S_p^{-1} * X - \frac{1}{2}(\bar{X}_1-\bar{X}_2)' * S_p^{-1}(\bar{X}_1+\bar{X}_2) < \ell n\left[\frac{C(1/2)}{C(2/1)} * \frac{P_2}{P_1}\right] \quad \mathbf{(33-3)}$$

فإذا أخذنا أية نقطة مثل x_0 وكانت تحقق العلاقة الأولى (3−23) فإننا ننسب x_0 إلى G_1 وإذا كان العكس ننسبها إلى G_2 .

ملاحظة: إن الحد الأول في العلاقة (3−32) ماهو إلا معادلة تابع خطي لـ X لذلك سنرمز له بـ Y، فيكون لدينا :

$$Y = (\bar{X}_1-\bar{X}_2)' * S_p^{-1} * X = \ell' * X \quad (34-3)$$

أما الحد الثاني فهو عبارة عن عدد حقيقي ناتج عن جداءات الأطراف التي فيه .

ملاحظة: يمكن أن نرمز للطرف الأيسر في العلاقة (3−32) بـ W كما فعل أندرسون *(Anderson)* فيكون لدينا :

$$W(X) = (\bar{X}_1-\bar{X}_2)' * S_p^{-1} * X - \frac{1}{2}(\bar{X}_1-\bar{X}_2)' * S_p^{-1} * (\bar{X}_1+X_2) =$$

$$W(X) = (\bar{X}_1-\bar{X}_2)' * S_p^{-1} * \left[X - \frac{1}{2}(\bar{X}_1+\bar{X}_2)\right] \quad (35-3)$$

وهي معادلة مستقيم ميله $\left[(\bar{X}_1-\bar{X}_2)'S_p^{-1}\right]$ ويمر من النقطة الوسطى $\widetilde{m} = \frac{1}{2}(\bar{X}_1+\bar{X}_2)$

وهذه العلاقة تسمى تابع التصنيف لأندرسون *(Anderson)* وهي تشير إلى أن المستوى *W(X)* يمر من نقطة القيمة الوسطى m. وبذلك فإن قاعدة التصنيف (3−32) و(3−33) تأخذ الشكل التالي:

$$R_1 : W(X) \geq \ell n \left[\left[\frac{C(1/2)}{C(2/2)} * \frac{P_2}{P_1}\right]\right] \qquad (36-3)$$

$$R_2 : W(X) < \ell n \left[\left[\frac{C(1/2)}{C(2/1)} * \frac{P_1}{P_2}\right]\right] \qquad (36\,a-3)$$

أما إذا كان الطرف الأيمن في (3−32): $\left[\frac{C(1/2)}{C(2/1)} * \frac{P_2}{P_1}\right] = 1$ فإن $\ell n \left[\frac{C(1/2)}{C(2/1)} * \frac{P_2}{P_1}\right] = 0$

وعندها تصبح قاعدة التصنيف كما يلي: إذا كان x_0 أي عنصر من المجتمع فإننا نصنفه كما يلي :

(37−3) إذا كان $W(x_0) \geq 0$ فإننا نصنف x_0 في المجموعة G_1

($37a$ −3) وإذا كان $W(x_0) < 0$ فإننا نصنف x_0 في المجموعة G_2

وهذا يتفق مع قاعدة المستقيم الفاصل بين المجموعتين، أي أنه إذا كان لدينا مجتمعان طبيعيان وكان لهما نفس مصفوفة التباين المشترك، فإن قاعدة المستقيم الفاصل بينهما تكافئ قاعدة القيمة المتوقعة للتكلفة ECM، وذلك عندما تكون فيها الاحتمالات المسبقة متساوية وتكون تكاليف التصنيف الخاطئ متساوية .

وبصورة عامة إذا كانت المتحولات طبيعية ومصفوفات التباينات متساوية، فإن مؤشر التصنيف W(X) المعرف في (3−35) يمكن أن يحسب لكل مشاهدة x_0، ثم يتم العمل على تصنيف المشاهدات بمقارنة قيم W(X)مع قيمة اللوغاريتم$\ell n \left[\frac{C(1/2)}{C(2/1)} * \frac{P_2}{P_1}\right]$ ويتم التصنيف كما في العلاقتين(3−36) و(3− $36a$) .

مثال 3−1: لنفترض أنه لدينا مجموعتين G_1 وG_2 ونريد تمييزهما بمتحولين طبيعيين X_1 وX_2، لذلك سحبنا منهما عينتين بحجمين $n_1 = n_2 = 3$ فحصلنا منهما على البيانات التالية:

$$XG_1 = \begin{bmatrix} 2 & 4 & 3 \\ 12 & 10 & 2 \end{bmatrix} \quad , \quad \bar{X}_1 = \begin{bmatrix} 3 \\ 10 \end{bmatrix} \quad , \quad S_1 = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$$

$$XG_2 = \begin{bmatrix} 5 & 3 & 4 \\ 7 & 9 & 5 \end{bmatrix} \quad , \quad \bar{X}_2 = \begin{bmatrix} 4 \\ 7 \end{bmatrix} \quad , \quad S_2 = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$$

وبما أن مصفوفتي التباينين $S_1 = S_2$ ، فإننا نحسب مصفوفة التباين المدمجة منهما بواسطة العلاقة (3−31) فنجد أن:

$$S_p = \frac{1}{3+3-2}\left[2\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} + 2\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}\right] = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix} = \tilde{V}$$

ثم نحسب مقلوبها S_p^{-1} فنجد أن:

$$S_p^{-1} = \frac{1}{4-1}\begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}$$

وبفرض أن الاحتماليين المسبقين للمجموعتين متساويان $(P_1 = P_2)$ وأن تكاليف التصنيف الخاطئ متساوية $(C(1/2) = C(2/1))$، فعندها سيكون لدينا $\ell n\left[\frac{C(1/2)}{C(2/1)} * \frac{P_2}{P_1}\right] = 0$ وهو الطرف الأيمن للعلاقة (3–32)، ثم نقوم بحساب معادلة المؤشر *W(X)* من العلاقة (3–35) فنجد أن:

$$W(X) = (\bar{X}_1 - \bar{X}_2)' * S_p^{-1} * \left[X - \frac{1}{2}(\bar{X}_1 + \bar{X}_2)\right]$$

$$W(X) = \left(\begin{bmatrix}3\\10\end{bmatrix} - \begin{bmatrix}4\\7\end{bmatrix}\right)' * \frac{1}{3}\begin{bmatrix}4 & 1\\1 & 1\end{bmatrix} * \left(\begin{bmatrix}X_1\\X_2\end{bmatrix} - \frac{1}{2}\left(\begin{bmatrix}3\\10\end{bmatrix} + \begin{bmatrix}4\\7\end{bmatrix}\right)\right)$$

$$W(X) = \left(\frac{1}{3}[-1\ \ ,3]\begin{bmatrix}4 & 1\\1 & 1\end{bmatrix} * \begin{bmatrix}X_1 - \frac{7}{2}\\ X_2 - \frac{17}{2}\end{bmatrix}\right)$$

$$W(X) = \frac{1}{3}[-1\ \ ,2]\begin{bmatrix}X_1 - \frac{7}{2}\\ X_2 - \frac{17}{2}\end{bmatrix} = \frac{1}{3}\left[\left(-X_1 + \frac{7}{2}\right) + (2X_2 - 17)\right]$$

وهي معادلة التابع التمييزي الفاصل بين المجموعتين : $W(X) = -\frac{1}{3}X_1 + \frac{2}{3}X_2 - \frac{27}{6}$

فإذا أخذنا المشاهدة الأولى $\begin{bmatrix}2\\12\end{bmatrix}$ من G_1 نجد أن *W(X)* المقابلة لها تساوي:

$$W_1 = -\frac{1}{3}2 + \frac{2}{3}12 - \frac{27}{6} = 8 - \frac{31}{6} > 0$$

أي أن المشاهدة $\begin{bmatrix}2\\12\end{bmatrix}$ ستصنف في المجموعة الأولى G_1 وهي أصلاً من G_1 وهو تصنيف صحيح .

وإذا أخذنا المشاهدة الثانية $\begin{bmatrix}4\\10\end{bmatrix}$ من G_1 نجد أن *W(X)* تساوي:

$$W_2 = -\frac{1}{3}4 + \frac{2}{3}10 - \frac{27}{6} = \frac{20}{3} - \frac{35}{6} > 0$$

أي أن المشاهدة $\begin{bmatrix}4\\10\end{bmatrix}$ ستصنف أيضاً في G_1 وهو تصنيف صحيح أيضاً .

وإذا أخذنا المشاهدة $\begin{bmatrix}3\\8\end{bmatrix}$ من G_1 نجد أن W تساوي :

$$W_3 = -\frac{1}{3}3 + \frac{2}{3}8 - \frac{27}{6} = \frac{16}{3} - \frac{33}{6} < 0$$

أي أن المشاهدة $\begin{bmatrix}3\\8\end{bmatrix}$ ستصنف في المجموعة G_2 وهي أصلاً من G_1 وهذا تصنيف خاطئ .

وإذا أخذنا المشاهدة الأولى $\begin{bmatrix}5\\7\end{bmatrix}$ من G_2 نجد أن *W(X)* تساوي:

$$W_4 = -\frac{1}{3}5 + \frac{2}{3}7 - \frac{27}{6} = \frac{14}{3} - \frac{37}{6} < 0$$

أي أن المشاهدة $\begin{bmatrix}5\\7\end{bmatrix}$ ستصنف في المجموعة G_2 وهي أصلاً من G_2 وهو تصنيف صحيح .

وإذا أخذنا المشاهدة الثانية $\begin{bmatrix}3\\9\end{bmatrix}$ من G_2 نجد أن *W(X)* تساوي:

$$W_5 = -\frac{1}{3}3 + \frac{2}{3}9 - \frac{27}{6} = 5 - \frac{37}{6} > 0$$

أي أن المشاهدة $\begin{bmatrix}3\\9\end{bmatrix}$ ستصنف في المجموعة G_1 وهي أصلاً من G_2 وهو تصنيف خاطئ .

وإذا أخذنا المشاهدة الثالثة $\begin{bmatrix}4\\5\end{bmatrix}$ من G_2 نجد أن *W(X)* تساوي:

$$W_6 = -\frac{1}{3}4 + \frac{2}{3}5 - \frac{27}{6} = \frac{10}{3} - \frac{35}{6} < 0$$

أي أن المشاهدة $\begin{bmatrix}3\\9\end{bmatrix}$ من G_2 ستصنف في المجموعة G_2 وهي أصلاً من G_2 وهو تصنيف صحيح .

وبذلك نحصل على جدول تقاطع التصنيف الجديد مع التوزيع الأصلي التالي :

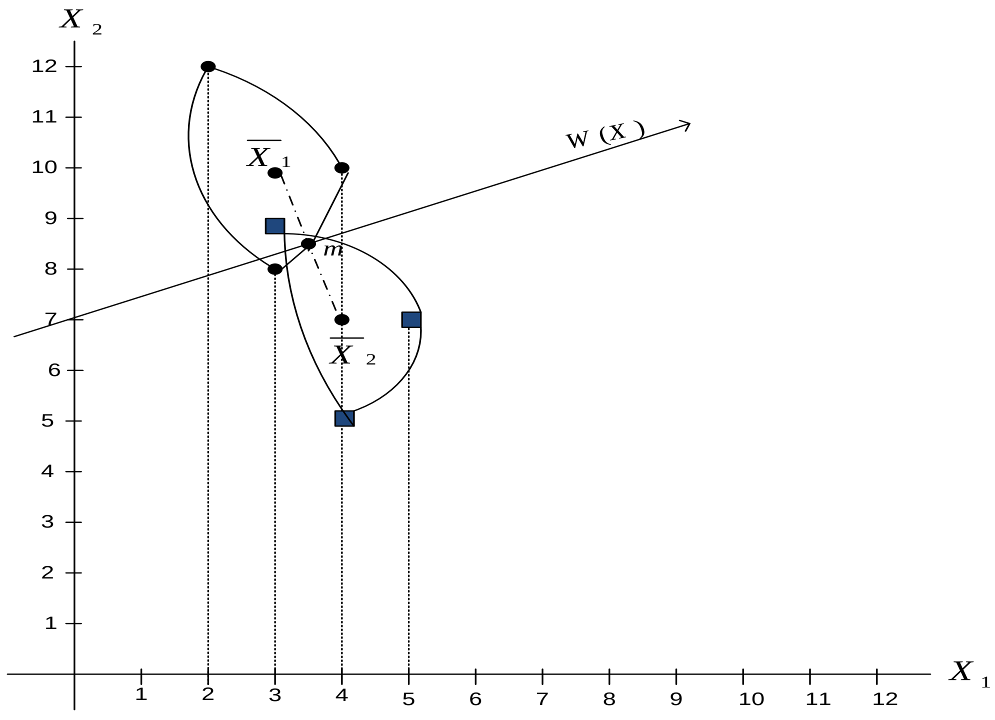
المجموع	التصنيف الجديد			
	G_2	G_1	المجموعة	التوزيع الاصلي
3	$n_{1m} = 1$	$n_{11} = 2$	G_1	
3	$n_{22} = 2$	$n_{2m} = 1$	G_2	

وبناءً على ذلك يمكننا حساب معدل الخطأ الظاهري *APEP* (*Apparent Error Rate*) وهو المعرف بالعلاقة التالية:

$$APEP = \frac{n_{1m} + n_{2m}}{n_1 + n_2} = \frac{1+1}{6} = 0.333 \qquad (38-3)$$

وهو احتمال كبير نسبياً، ويمكن أن يكون أكبر من ذلك لو إننا لم نعتبر الاحتمال المسبق والتكاليف متساوية، ويعود ذلك إلى صغر حجمي العينتين n_2 و n_1، ويمكن أن تقل قيمة هذا الاحتمال عندما نجعل حجمي العينتين كبيرين .

ويمكننا رسم هذه النقاط على الشكل البياني كما يلي:



الشكل (3-4): توزيع عناصر المجموعتين G_1 و G_2 حيث أن

● عناصر المجموعة G_1 ■ عناصر المجموعة G_2

ومن الشكل نلاحظ أن المستقيم الفاصل بين عناصر هاتين المجموعتين يقسمها إلى مجموعتين جديدتين كما هو مبين في الجدول السابق. ونحصل على معادلة المستقيم الفاصل في المستوى $X_1 0 X_2$ بوضع $W = 0$ فنجد أن :

$$\frac{2}{3}X_2 = \frac{1}{3}X_1 + \frac{27}{6}$$
$$X_2 = \frac{1}{2}X_1 + \frac{27}{4}$$

وهو يمر من النقطة $\left(\frac{27}{4}, 0\right)$ ومن القيمة الوسطى $\left(\frac{7}{2} , \frac{17}{2}\right)$، ولكنه لا يتعامد مع القطعة المستقيمة الواصلة بين المركزين $\bar{X}_1$و $\bar{X}_2$ لأن ميله $m' = \frac{1}{2}$ لا يساوي $\left(-\frac{1}{m}\right)$حيث أن $\left(m = \frac{10-7}{3-4} = -3\right)$

- **الحالة الثانية:** وهي التي يكون فيها $V_1 \neq V_2$

وهذه الحالة هي الحالة العامة التي تفترض أن مصفوفتي التباين المشترك للمتحولين $\left(X_1\text{و } X_2\right)$ في المجموعتين $\left(G_1\text{و } G_2\right)$ غير متساويين .

وهنا يمكننا إجراء نفس المعالجات التي أجريناها في الحالة $V_1 = V_2$، ثم نقوم بحساب نسبة التوزيعين $f_1(x)/f_2(x)$ ثم حساب لوغاريتم هذه النسبة من العلاقة :

$$\ell n\frac{f_1(x)}{f_2(x)} = \ell n f_1(x) - \ell n f_2(x) \qquad (38-3)$$

وبعد تعويض معادلتي $f_1(x)$ و$f_2(x)$ من العلاقتين (3–20) و(3–21) وبإجراء بعض الإصلاحات نحصل على منطقتي التصنيف المعرفين كما يلي:

$$R_1: -\frac{1}{2}X'\left(\boldsymbol{V_1^{-1}} - \boldsymbol{V_2^{-1}}\right) * \boldsymbol{X} + \left(\boldsymbol{\mu_1' V_1^{-1}} - \boldsymbol{\mu_2' V_2^{-1}}\right) * \boldsymbol{X} - \boldsymbol{K} \geq \ell n\left[\frac{C(1/2)}{C(2/1)}\right]\left[\frac{P_2}{P_1}\right] \quad (39-3)$$
$$R_2: -\frac{1}{2}X'\left(\boldsymbol{V_1^{-1}} - \boldsymbol{V_2^{-1}}\right) * \boldsymbol{X} + \left(\boldsymbol{\mu_1' V_1^{-1}} - \boldsymbol{\mu_2' V_2^{-1}}\right) * \boldsymbol{X} - \boldsymbol{K} < \ell n\left[\frac{C(1/2)}{C(2/1)}\right]\left[\frac{P_2}{P_1}\right] \quad (40-3)$$

حيث أن: K هو عدد يحسب من العلاقة التالية:

$$K = \frac{1}{2}\ell n\left(\frac{|V_1|}{|V_2|}\right) + \frac{1}{2}\left(\boldsymbol{\mu_1'} V_1^{-1}\boldsymbol{\mu_1} - \boldsymbol{\mu_2'} V_2^{-1}\boldsymbol{\mu_2}\right) \qquad (41-3)$$

وهنا نلاحظ أن الحد الأول $\left(-\frac{1}{2}X'(V_1^{-1} - V_2^{-1}) * X\right)$ في العلاقة (3–39) يعطي بدلالة تابع تربيعي لـ X (من الدرجة الثانية لـ X)، أي أن الفاصل سيكون على شكل منحني من الدرجة الثانية لـ X. كما نلاحظ أنه عندما: $V_1 = V_2$ فإن هذا الحد يختفي وتعود العلاقة (3–39) إلى العلاقة (3–32) ذات التابع الخطي.

وعندما تكون المعالم $\boldsymbol{\mu_1}$و $\boldsymbol{\mu_2}$ و V_1و V_2 غير معلومة في المجتمع، فإننا نستبدلها بتقديراتها غير المتحيزة المحسوبة من العينتين العشوائيتين $\bar{X}_1$و $\bar{X}_2$ و S_1و S_2 على الترتيب، وعند تعويض ذلك في العلاقات (3–39) و(3–40) و(3–41) نحصل على قاعدة تصنيف أي عنصر من المجتمع مثل X_0

في إحدى المجموعتين كما يلي:

نصنف X_0 في المجموعة G_1 إذا تحققت المتراجحة التالية:

$$-\frac{1}{2}X_0'(S_1'-S_2')X_0+(\bar{X}_1'S_1^{-1}-\bar{X}_2'S_2^{-1})X_0-K\geq ln\left[\frac{C(1/2)}{C(2/1)}\right]\left[\frac{P_2}{P_1}\right] \qquad (\mathbf{42-3})$$

ونصنفه في المجموعة G_2 إذا كانت المتراجحة (3−42) غير محققة .

علماً بأن K تحسب أو تقدر من العلاقة التالية:

$$K=\frac{1}{2}\ ln\left(\frac{|V_1|}{|V_2|}\right)+\frac{1}{2}\left(\overline{X}_1^{'}S_1^{-1}\bar{x}-\overline{X}_2^{'}S_2^{-1}\bar{x}\right) \qquad (43-3)$$

ملاحظة: كتعقيب على هذه الحالة نلاحظ أن التصنيف وفق التابع التربيعي يكون أكثر تعقيداً في الفضاءات العليا (أكثر من متحولين) ويمكن أن يقودنا إلى بعض النتائج الغريبة .

وإذا كانت البيانات غير خاضعة للتوزيع الطبيعي المتعدد فهناك رأيان ممكنان لمعالجة ذلك هما:

<u>الأول:</u> يمكن أن نقوم بتحويل البيانات غير الطبيعية إلى بيانات قريبة من الطبيعة، ثم اختبار تساوي مصفوفتي التباين المشترك (قبل وبعد التحويل)، ثم تحديد أي الحالتين أفضل للتصنيف. حالة التابع الخطي أم حالة التابع التربيعي .

وإن أهم المعادلات المستخدمة في التحويل إلى طبيعي أو شبه طبيعي هي:

1−للمتحول الرقمي y الموجب يمكن أن نستخدم أحد التحويلين $\sqrt{y}$ أو $\log y$.

2−للمتحول النسبي $\tilde{P}$ نستخدم التحويل المنطقي: $\mathrm{logit}\tilde{P}=\frac{1}{2}\log\left(\frac{\tilde{P}}{1-\tilde{P}}\right)$

3−لمعاملات الارتباط r نستخدم تحويل فيشر: $\mathrm{Z(r)}=\frac{1}{2}\log\left(\frac{1+\mathrm{r}}{1-\mathrm{r}}\right)$

الثاني: يمكننا استخدام التابع الخطي أو التربيعي بدون قلق أو اهتمام بخصائص المجتمع الأصلي. ولقد قام فيشر بذلك، دون أن يعتمد على خصائص المجتمع الأصلي. وانطلق من فكرة تعظيم المسافة النسبية بين متوسطي مجموعتين طبيعيتين أو غير طبيعيتين. رغم أن نتائج بعض الحالات غير الطبيعية كانت غريبة .

لذلك يصبح من الواجب علينا دائماً أن نقوم باختبار نتائج أي أسلوب يمكن أن يستخدم للتصنيف، وذلك من أجل إنشاء التابع التصنيفي المناسب.

ومن الناحية المثالية يفضل أن يكون حجم البيانات كافياً ليبرهن على صحة العينات (التدريبية) والعينات (التصديقية). حيث أن العينات (التدريبية) تستخدم لإنشاء توابع التصنيف ويجب أن تكون حجومها كبيرة أو كافية لتؤمن ثقة كافية في التوابع المستخدمة. أما العينات (التصديقية) فهي يستخدم لتقييم أداء تلك التوابع. ويمكن أن تكون حجومها صغيرة نسبياً لتؤكد صحة التوابع التصنيفية المستخدمة .

5-3 تقييم توابع التصنيف: [Johnston P. 494 بتصرف وإضافة]

إن أهم وسيلة للحكم على أداء أي أسلوب للتصنيف هو حساب (معدل الخطأ) فيه، أو حساب اجمالي احتمالات التصنيف الخاطئ TPM الذي يعطي بالعلاقة (2-12) التالية:

$$TPM = P_1 \int_{R_2} f_1(x)dx + P_2 \int_{R_1} f_2(x)dx \qquad (44-3)$$

إن أصغر قيمة لـ TPM التي يمكن الحصول عليها باختيار وتحديد المنطقتين R_1 و R_2 (أو تحديد الحد الفاصل بينها)، تسمى بمعدل الخطأ المثالي *OER (Optimum error rate)* ونكتب ذلك كمايلي:

$$OER = Min\left[P_1 \int_{R_2} f_1(x)dx + P_2 \int_{R_1} f_2(x)dx\right] \qquad (45-3)$$

أي أن OER هو معدل الخطأ الأصغر لتابع التصنيف .

ولتحديد المنطقتين R_1 و R_2 نفترض أن تكاليف التصنيف الخاطئ متساوية، ونستخدم العلاقتين (3-22) و(3-23) فنحصل على ما يلي:

$$R_1: \frac{f_1(x)}{f_2(x)} \geq \frac{P_2}{P_1} \qquad\qquad R_2: \frac{f_1(x)}{f_2(x)} < \frac{P_2}{P_1} \qquad (46-3)$$

مثال (3-2): يطلب منا أن نوضح معنى معدل الخطأ المثالي OER لمجتمع مؤلف من مجموعتين: G_1 و G_2 متساويتين الحجم، أي أن نسبتيهما $\left(p_1 = p_2 = \frac{1}{2}\right)$، وتتأثران بمتحولين X_1 و X_2 خاضعين فيهما للتوزيعين الطبيعيين $f_1(x)$ و $f_2(x)$ المعرفتين بالعلاقتين (3-20) و(3-21)، وإن تكاليف التصنيف الخاطئ متساوية، وإن $(V_1 = V_2 = V)$.

نلاحظ أن أصغر قيمة لـ ECM وأصغر قيمة لـ TPM لأي تابع تمييزي تتطابقان عندما يكون $C(1/2) = C(2/1)$ وعندما $p_1 = p_2$ ، ولأنه عندها يكون الطرف الأيمن من العلاقة (3-29) معدوماً لأن:

$$\ell n\left[\frac{C(1/2)}{C(2/1)}\right]\left[\frac{P_2}{P_1}\right] = \ ln(1) = 0$$

ومنها نستنتج أن المنطقتين R_1 و R_2 تتحددان في هذه الحالة من العلاقتين:

$$R_1: (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * V^{-1} * X - \frac{1}{2}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * V^{-1} * (\boldsymbol{\mu_1} + \boldsymbol{\mu_2}) \geq 0$$

$$R_2: (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * V^{-1} * X - \frac{1}{2}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * V^{-1} * (\boldsymbol{\mu_1} + \boldsymbol{\mu_2}) < 0$$

وإذا رمزنا للحد الأول فيهما بالرمز y يكون لدينا:

$$y = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * V^{-1} * X = \ell' * X \qquad (47-3)$$

حيث رمزنا بـ ℓ' للمقدار $\ell' = (\mu_1 - \mu_2)' * V^{-1}$ فيكون لدينا :

$$\ell = V^{-1}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) \qquad (48-3)$$

وبعدها نجد أن المنطقتين R_1 و R_2 تصبحان تابعين لـ y، ويتم تحديدهما من العلاقتين:

$$R_1(y): y \geq \frac{1}{2}(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})' * V^{-1} * (\boldsymbol{\mu_1}+\boldsymbol{\mu_2}) \qquad (48-3)$$

$$R_2(y): y < \frac{1}{2}(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})' * V^{-1} * (\boldsymbol{\mu_1}+\boldsymbol{\mu_2}) \qquad (49-3)$$

وبما أن y هو تركيب خطي لمتحولات طبيعية عشوائية، فإن التوزيع الاحتمالي لـ y هو توزيع طبيعي بمتحول واحد y في المجموعتين، ونرمز له بـ $f_1(y)$ في G_1 وبـ $f_2(y)$ في G_2 .

وإن متوسطي y في المجموعتين يساويان:

$$\mu_{1y} = \ell' * \boldsymbol{\mu_1} = (\boldsymbol{\mu_1}-\boldsymbol{\mu_2})' * V^{-1} * \mu_1 \qquad (50-3)$$

$$\mu_{2y} = \ell' * \boldsymbol{\mu_2} = (\boldsymbol{\mu_1}-\boldsymbol{\mu_2})' * V^{-1} * \mu_2 \qquad (51-3)$$

وإن تباينه في كل من المجموعتين يساوي: $\sigma_{y_1}^2 = \sigma_{y_2}^2 = \sigma_y^2$ ويحسب من العلاقة:

$$\sigma_y^2 = \ell' * V * \ell = (\boldsymbol{\mu_1}-\boldsymbol{\mu_2})' * V^{-1} * V * V^{-1} * (\boldsymbol{\mu_1}-\boldsymbol{\mu_2})$$

$$\sigma_y^2 = (\boldsymbol{\mu_1}-\boldsymbol{\mu_2})' * V^{-1} * (\boldsymbol{\mu_1}-\boldsymbol{\mu_2}) = \Delta^2 \qquad (51-3)$$

وبذلك تأخذ علاقة TPM الشكل التالي:

$$TPM(y) = P_1 \int_{R_2} f_1(y)dy + P_2 \int_{R_1} f_2(y)dy \qquad (52-3)$$

وبناءً على العلاقة (3–49) نجد أن التكامل الأول يساوي:

$$P(2/1) = \int_{R_2} f_1(y)dy = P\left[y < \frac{1}{2}(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})' * V^{-1} * (\boldsymbol{\mu_1}+\boldsymbol{\mu_2})\right]$$

والآن نحول y إلى متحول معياري Z بإجراء التحويل $\left(Z = \frac{y-\mu_{1y}}{\sigma_y}\right)$ فنجد أن (3–50) تعطينا أن:

$$\mathrm{P}(2/1) = P\left[\frac{y-\mu_{1y}}{\sigma_y} < \frac{\frac{1}{2}(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})' * V^{-1} * (\boldsymbol{\mu_1}+\boldsymbol{\mu_2}) - (\boldsymbol{\mu_1}-\boldsymbol{\mu_2})' * V^{-1} * \boldsymbol{\mu_1}}{\sigma_y}\right]$$

وبعد الإصلاح واعتماداً على (3–16) نجد أن:

$$\mathrm{P}(2/1) = P\left[Z < \frac{-\frac{1}{2}(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})' * V^{-1} * (\boldsymbol{\mu_1}-\boldsymbol{\mu_2})}{\Delta}\right]$$

$$P(2/1) = \int_{R_2} f_1(y)dy = P\left(Z < \frac{-\frac{1}{2}\Delta^2}{\Delta}\right) = P\left(Z < \frac{-\Delta}{2}\right) = \phi\left(\frac{-\Delta}{2}\right) \qquad (53-3)$$

حيث $\phi\left(\frac{-\Delta}{2}\right)$ هي قيمة تابع التوزيع الطبيعي المعياري عند النقطة $\left(\frac{-\Delta}{2}\right)$.

وبطريقة مشابهة نجد أن التكامل الثاني يساوي:

$$P(1/2) = \int_{R_1} f_2(y)dy = P\left[y \geq \frac{1}{2}(\boldsymbol{\mu_1}-\boldsymbol{\mu_2})' * V^{-1} * (\boldsymbol{\mu_1}+\boldsymbol{\mu_2})\right]$$

$$P(1/2) = P\left[Z \geq \left(\frac{\Delta}{2}\right)\right] = 1 - \phi\left(\frac{\Delta}{2}\right) = \phi\left(\frac{-\Delta}{2}\right) \qquad (\mathbf{54-3})$$

ومن (3−53) و(3−54) نجد أن الاحتمال الاجمالي للتصنيف الخاطئ TPM في (3−52) يساوي:

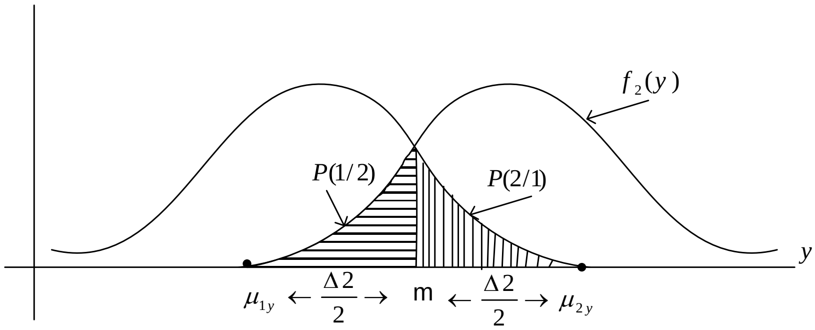
$$TPM = P_1\phi\left(\frac{-\Delta}{2}\right) + P_2\phi\left(\frac{-\Delta}{2}\right) \qquad \mathbf{(55-3)}$$

وأخيراً نجد أن معدل الخطأ المثالي OER يساوي:

$$OER = Minmum\, TPM = P_1 * \phi\left(\frac{-\Delta}{2}\right) + P_2 * \phi\left(\frac{-\Delta}{2}\right)$$

$$OER = \phi\left(\frac{-\Delta}{2}\right)(P_1 + P_2) = \phi\left(\frac{-\Delta}{2}\right) \qquad \mathbf{(56-3)}$$

وإن احتمالات هذا التصنيف الخاطئ مبينة على الشكل التالي:



الشكل (3−5) احتمالات التصنيف الخاطئ حسب y

حيث أن:

$$\Delta^2 = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * V^{-1} * (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$$

مثال تطبيقي: لنفترض أن:

$$\Delta^2 = \left(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}\right)' * V^{-1} * \left(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}\right) = 2.56$$

فعندها يكون لدينا: $\Delta = \sqrt{2.56} = 1.6$

ومن جداول تابع التوزيع الطبيعي المعياري نجد أن:

$$Min\, mum\, TPM = OER = \phi\left(\frac{-1.6}{2}\right) = 0.2119$$

وهو معدل غير قليل ويشير إلى توزيع خاطئ لبعض عناصر المجتمع على المجموعة الأولى G_1 أو على المجموعة الثانية G_2 .

ملاحظة: عندما تكون المعالم μ_1 و μ_2 و V غير معلومة. فإننا نستبدلها بتقديراتها من العينة بـ $\bar{X}_1$ و $\bar{X}_2$ و S_p . حيث أن S_p هو التباين المدمج من التباينين S_2 و S_1 وفق العلاقة (3−31) .

3−6 طريقة فيشر Fisher لمتحولين ومجموعتين: [Johnston, A. R. P.470 بتصرف وإضافة]

قبل استعراض هذه الطريقة نقدم بعض خواص المتحولات العشوائية ونختصرها بما يلي:

1−إذا كان a عدداً ثابتاً فإن التوقع الرياضي لـ (aX_i) يساوي :

$$E(aX_i) = a * E(X_i) = a\boldsymbol{\mu_i} \qquad \left(\text{حيث } \mu_i \text{ هو توقع } X_i\right) \qquad (57-3)$$

وإن تباين (aX_i) يساوي :

$$var(aX_i) = E(aX_i - a\boldsymbol{\mu_i})^2 = a^2E(X_i - \boldsymbol{\mu_i})^2 = a^2\sigma_{ii} \qquad (58-3)$$

2-إذا كان a و b عددين ثابتين فإن التباين المشترك للمتحولين aX_1 و bX_2 يساوي:

$$cov(aX_1\,, bX_2) = E(aX_1 - a\boldsymbol{\mu_1})(bX_2 - a\boldsymbol{\mu_2}) \qquad (59-3)$$
$$= abE(X_1 - \boldsymbol{\mu_1})(X_2 - \boldsymbol{\mu_2}) = ab\sigma_{12}$$

3-إذا كان a و b عددان ثابتين وكان $X_1\,, X_1$ متحولين عشوائيين فإن التوقع الرياضي لأي تركيب خطي لهما مثل $Y = aX_1 + bX_2$ وتباينه يساويان:

$$E(Y) = E\,(aX_1 + bX_2) = aE(X_1) + bE(X_2) = a\boldsymbol{\mu_1} + b\boldsymbol{\mu_2} \qquad (60-3)$$
$$var(Y) = var\,(aX_1 + bX_2) = E[(aX_1 + bX_2) - (a\boldsymbol{\mu_1} + b\boldsymbol{\mu_2})]^2$$
$$= E[a(X_1 - \boldsymbol{\mu_1}) + b(X_2 - \boldsymbol{\mu_2})]^2$$
$$= E[a^2(X_1 - \boldsymbol{\mu_1})^2 + b^2(X_2 - \boldsymbol{\mu_2})^2 * 2ab(X_1 - \boldsymbol{\mu_1})(X_2 - \boldsymbol{\mu_2})]$$
$$var(y) = a^2var(X_1) + b^2var(X_2) + 2ab\,cov(X_1\,, X_2)$$

أي أن:

$$var(aX_1 + bX_2) = a^2\sigma_{11} + b^2\sigma_{22} + 2ab\sigma_{12} \qquad (61-3)$$

وإذا رمزنا لشعاع العددين a و b بالرمز C فيكون لدينا:

$$C = \begin{bmatrix} a \\ b \end{bmatrix} \qquad C' = [a\;, b]$$

وعندها يمكننا كتابة التركيب الخطي السابق كما يلي:

$$Y = aX_1 + bX_2 = [a\;, b] * \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = C' * X \qquad (62-3)$$

وبذلك نجد أن توقعه يساوي:

$$EY = E(C' * X) = [a\;, b] * \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = C' * \boldsymbol{\mu} \qquad (63-3)$$

وإن تباينه يمكن أن يكتب بدلالة مصفوفة التباين المشترك V للمتحولين X_1 و X_2 كما يلي:

$$var(Y) = var\,(aX_1 + bX_2) = var[C'X] \qquad (64-3)$$
$$= C'var(X)C = C' * V * C$$

وهذا يعني أن:

$$var\,[aX_1 + bX_2] = [a\;, b]\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & 22 \end{bmatrix} * \begin{bmatrix} a \\ b \end{bmatrix}$$

وهو يساوي:

$$var\,[aX_1 + bX_2] = a^2\sigma_{11} + b^2\sigma_{22} + 2ab\sigma_{12} \qquad (65-3)$$

وانطلاقاً من الأفكار السابقة اقترح (فيشر) أن يتم تحويل التحليل التمييزي لتأثير المتحولين X_1 و X_2 على المجموعتين G_1 و G_2 إلى تحليل بسيط بمتحول واحد Y، ثم العمل على جعل قيم Y المشتقة من مشاهدات المجموعتين G_1 و G_2 بعيدة عن بعضها بأكبر قدر ممكن .

ولتحقيق هذه الفكرة اقترح (فيشر) أن يتم تشكيل متحول جديد Y بواسطة تركيب خطي للمتحولين X_1 و X_2 كما يلي:

$$Y = \ell_1X_1 + \ell_2X_2 = [\ell_1\,, \ell_2\,] \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \ell' * X \qquad (66-3)$$

وعند تعويض قيم X_1 وX_2 من كل مجموعة على حدة في هذا التركيب الخطي، نحصل على القيم العددية لـ Y المقابلة لمشاهدات كل مجموعة، وبالتالي نحصل على مجموعتين لقيمY نرمز لهما بما يلي:

$Y_1: y_{11}\ \ y_{12}\ \ y_{13} \dots . y_{1n_1}$ (للمجموعة G_1)

$Y_2: y_{21}\ \ y_{22}\ \ y_{23} \dots . y_{2n_2}$ (للمجموعة G_2)

ومنها يمكننا حساب متوسطي Y في المجموعتين من العلاقتين :

$$\mu_{1y} = \frac{1}{n_1}\sum_{i=1}^{n_1} y_{1i} \qquad (67-3)$$

$$\mu_{2y} = \frac{1}{n_2}\sum_{i=1}^{n_2} y_{2i} \qquad (68-3)$$

ومن جهة أخرى نجد أن المتوسطين μ_{1y} وμ_{2y} يمكن أن يحسبان بدلالة توقعي X_1 وX_2 في المجموعتين. فإذا افترضنا أن :

$$\mu_1 = E(X/G_1) = \begin{bmatrix}\boldsymbol{\mu_{11}}\\ \boldsymbol{\mu_{12}}\end{bmatrix} \qquad \mu_2 = E(X/G_2) = \begin{bmatrix}\boldsymbol{\mu_{21}}\\ \boldsymbol{\mu_{22}}\end{bmatrix} \qquad (69-3)$$

فعندها نجد أن:

$$\mu_{1y} = E(Y/G_1) = E(\ell' X/G_1) = \ell' * \boldsymbol{\mu_1} =$$
$$= (\ell_1 , \ell_2)\begin{bmatrix}\boldsymbol{\mu_{11}}\\ \boldsymbol{\mu_{12}}\end{bmatrix} = \ell_1\boldsymbol{\mu_{11}} + \ell_2\boldsymbol{\mu_{12}} \qquad (69-3)$$

$$\mu_{2y} = E(Y/G_2) = E(\ell' X/G_2) = \ell' * \boldsymbol{\mu_2} =$$
$$= (\ell_1 , \ell_2)\begin{bmatrix}\boldsymbol{\mu_{21}}\\ \boldsymbol{\mu_{22}}\end{bmatrix} = \ell_1\boldsymbol{\mu_{21}} + \ell_2\boldsymbol{\mu_{22}} \qquad (70-3)$$

ولحساب تباين Y بدلالة مصفوفة التباينات المشتركة لـ X_1 وX_2 نجد أن:

$$V_1 = E[(X - \boldsymbol{\mu_1})(X - \boldsymbol{\mu_1})'] \qquad (71-3)$$

$$V_2 = E[(X - \boldsymbol{\mu_2})(X - \boldsymbol{\mu_2})'] \qquad (72-3)$$

والآن لنفترض أن هاتين المصفوفتين متساويتان، أي نفترض أن $V_1 = V_2 = V$، فنجد أن تباين Y يساوي حسب (3−64) مايلي:

$$\sigma_y^2 = vor(Y) = vor(\ell' X) = \ell' * vor(X) * \ell$$

ومنها نجد أن :

$$\sigma_y^2 = \ell' * V * \ell \qquad (73-3)$$

وهكذا نصل إلى مكونات فكرة (فيشر) التي صاغها على شكل علاقة رياضية سميت (معيار فيشر)، وهي تهدف إلى جعل مربع المسافة النسبية بين متوسطي قيم Y في المجموعتين G_1 وG_2 أكبر مايمكن. ولقد صاغ (فيشر) معياره كما يلي:

$$\mathrm{F} = \frac{\left(\boldsymbol{\mu_{1y}} - \boldsymbol{\mu_{2y}}\right)^2}{\sigma_y^2} = \frac{(\ell'\boldsymbol{\mu_1} - \ell'\boldsymbol{\mu_2})^2}{\sigma_y^2} = \frac{[\ell'(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})]^2}{\sigma_y^2} = \frac{(\ell' * \boldsymbol{d})^2}{\ell' * V * \ell} \qquad (74-3)$$

حيث أن: $\boldsymbol{d}$ هو شعاع الفرق بين المتوسطين $\boldsymbol{d} = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$ ، والمطلوب الآن تحديد الأعداد في $\ell' = (\ell_1 , \ell_2)$ أو $\ell = \begin{pmatrix} \ell_1 \\ \ell_2 \end{pmatrix}$ التي تجعل قيمة المعيار F أكبر ما يمكن .

ولتحديد مركبات ℓ نستند إلى مسألة تعظيم المقدار الأخير الواردة في الملحق (2)، وحسب متراجحة (كوشي شوارتز) (2−70) في الملحق (2) وما يليها فنجد أنها تعطينا أن أكبر قيمة للمقدار(3−74) تساوي :

$$Max\left[\frac{\left(\ell' d\right)^2}{\ell' * V * \ell}\right] = \boldsymbol{d}' * V^{-1} * \boldsymbol{d} \qquad (75-3)$$

وإن هذه القيمة الكبرى تحصل حسب العلاقة (*2- 70a*) في الملحق (2) عندما تأخذ ℓ القيمة التالية:

$$\ell = C * V^{-1} * \boldsymbol{d} \qquad (76-3)$$

حيث C عامل تناسب لا يساوي الصفر $C \neq 0$

وبتعويض قيمة $\boldsymbol{d}$ التي تساوي : $\boldsymbol{d} = \boldsymbol{\mu_1} - \boldsymbol{\mu_2}$ نحصل على أن الحل المناسب ℓ لتعظيم (3−75) هو الحل التالي:

$$\ell = C * V^{-1} * (\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) \qquad (77-3)$$

ومنه نحصل على لا نهاية من الحلول لـ ℓ تتناسب مع الجداء $V^{-1} * (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$.

وبوضع *C=1* وملاحظة أن V^{-1} متناظرة نحصل على الحل الخاص ℓ ومنقوله ℓ' التالي:

$$\ell = V^{-1}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) \implies \ell' = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) * V^{-1} \qquad (78-3)$$

وبتعويض ذلك الحل في تابع (فيشر) المعرف في (3−66)، يصبح كما يلي:

$$\tilde{Y} = \ell' * X = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * V^{-1} * X \qquad (79-3)$$

وهي علاقة شبيهة بالعلاقة (3−34) السابقة، وهو عبارة عن مستقيم يمر من مبدأ الاحداثيات ويتم اسقاط النقاط عليه وحساب المتوسطين $\boldsymbol{\mu_1}$ و $\boldsymbol{\mu_2}$ والقيمة الوسطى m .

ولاستخدام التابع (3−79) في عمليات تصنيف المشاهدات على المجموعتين G_1 و G_2، نقوم بحساب القيمة الوسطى للمتوسطين $\boldsymbol{\mu_{1y}}$ و $\boldsymbol{\mu_{2y}}$ فنحصل على أنها تساوي:

$$m = \frac{1}{2}\left(\boldsymbol{\mu_{1y}} + \boldsymbol{\mu_{2y}}\right) = \frac{1}{2}\left(\ell'\boldsymbol{\mu_1} + \ell'\mu_2\right) = \frac{1}{2}\ell'(\boldsymbol{\mu_1} + \boldsymbol{\mu_2}) \qquad (80-3)$$

وبتعويض ℓ' من (3−78) نحصل على أن القيمة الوسطى تساوي:

$$m = \frac{1}{2}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * V^{-1}(\boldsymbol{\mu_1} + \boldsymbol{\mu_2}) \qquad (81-3)$$

والآن لنفترض أنه لدينا مشاهدة عشوائية مثل $X_0 = \begin{bmatrix} x_{01} \\ x_{02} \end{bmatrix}$ من المجتمع المدروس ونريد تحديد المجموعة التي ستنتمي إليها هذه المشاهدة .

لذلك نقوم بتعويض إحداثيات هذه المشاهدة في العلاقة (3−79) فنحصل منها على قيمة محددة $\tilde{y}_0$

$$\tilde{y}_0 = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * V^{-1}X_0 = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})'V^{-1}\begin{bmatrix} x_{01} \\ x_{02} \end{bmatrix} \qquad (82-3)$$

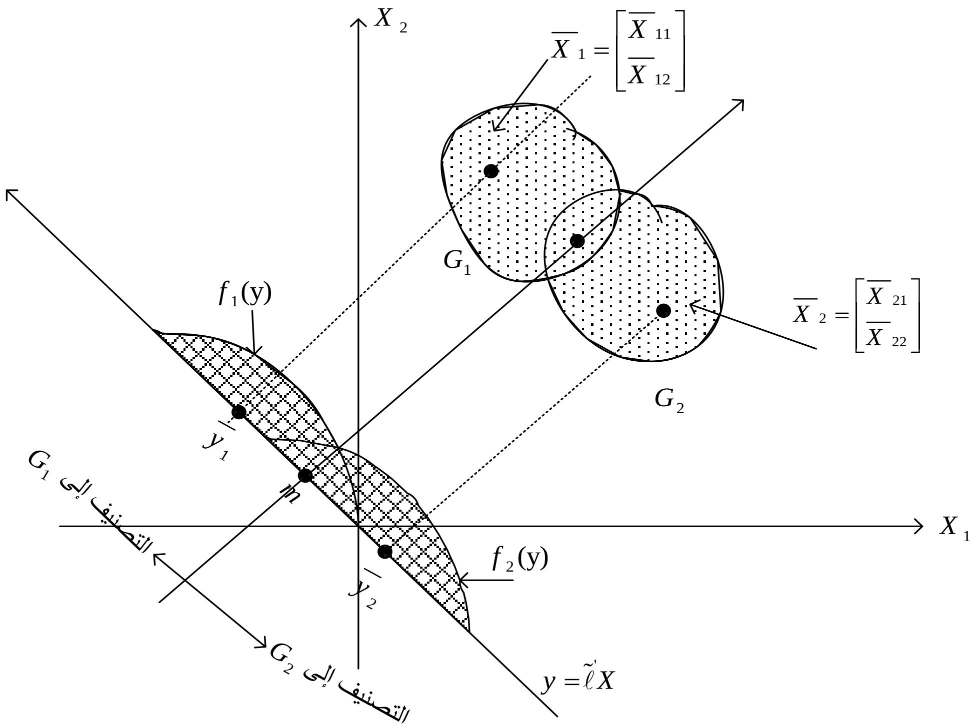
ثم نقوم بمقارنة هذه القيمة $\tilde{y}_0$ مع القيمة الوسطى m ونتخذ قرار التصنيف كمايلي:

$$\begin{cases} \text{إذا كانت } \tilde{y}_0 \geq m \text{ فإننا نصنف } X_0 \text{ في المجموعة } G_1 \\ \text{وإذا كانت } \tilde{y}_0 < m \text{ فإننا نصنف } X_0 \text{ في المجموعة } G_2 \end{cases} \quad (83-3)$$

وبعبارة أخرى يمكننا أن نصنف العناصر حسب القاعدة التالية :

$$\begin{cases} \text{إذا كانت } (\tilde{y}_0 - m) \geq 0 \text{ فإننا نصنف } X_0 \text{ في المجموعة } G_1 \\ \text{وإذا كانت } (\tilde{y}_0 - m) < 0 \text{ فإننا نصنف } X_0 \text{ في المجموعة } G_2 \end{cases} \quad (84-3)$$

والشكل البياني التالي يوضح ذلك :



الشكل (3-6): التصنيف حسب معيار (فيشر)

ملاحظة إذا كانت معالم المجتمع μ_1 و μ_2 و V مجهولة فإننا نقوم بتقدير الكميات μ_1 و μ_2 و V وحساب ℓ و m من خلال معلومات العينات المسحوبة عشوائياً من المجموعتين G_1 و G_2 .

والآن لنفترض إننا سحبنا عينة عشوائية من المجموعة الأولى بحجم n_1 وسحبنا عينة عشوائية أخرى من المجموعة الثانية بحجم n_2 وحصلنا منهما على مصفوفتي القياسات التالية:

المتوسط	قيم عناصر العينة الثانية n_2	المتحولات
$\bar{x}_{12}$	x_{121} x_{122} x_{123} ... x_{12n_2}	X_1
$\bar{x}_{22}$	x_{221} x_{222} x_{223} ... x_{22n_2}	X_2

المتوسط	قيم عناصر العينة الثانية n_1	المتحولات
$\bar{x}_{11}$	x_{111} x_{112} x_{113} ... x_{11n_1}	X_1
$\bar{x}_{21}$	x_{211} x_{212} $2x_{113}$ ... x_{21n_1}	X_2

ومنهما نحسب شعاعي المتوسطين فنجد من العلاقة $\bar{x} = \frac{1}{n}\sum x_i$ أن :

$$\bar{X}(G_1) = \bar{X}_1 = \begin{bmatrix} \bar{x}_{11} \\ \bar{x}_{21} \end{bmatrix} \qquad \bar{X}(G_2) = \bar{X}_2 = \begin{bmatrix} \bar{x}_{12} \\ \bar{x}_{22} \end{bmatrix} \qquad (85-3)$$

نعلم أن هذين المتوسطين هما تقديران غير متحيزين للتوقعين $\boldsymbol{\mu_1}$ و $\boldsymbol{\mu_2}$ في المجموعتين G_1 و G_2، اللتين يتألف منهما المجتمع المدروس، ونكتب ذلك كما يلي:

$$\tilde{\boldsymbol{\mu}}_1 = \bar{X}_1 \qquad \tilde{\boldsymbol{\mu}}_2 = \bar{X}_2 \qquad (86-3)$$

ولإيجاد تقدير غير متحيز لمصفوفة التباين المشترك V، نقوم بحساب مصفوفتي التباين المشترك للمتحولين X_1 و X_2 في كل مجموعة على حدة من العلاقتين التاليتين:

$$S_1 = \frac{1}{n_1-1}\sum_{j=1}^{n_1}(X_{1j}-\bar{X}_1)(X_{1j}-\bar{X}_1)' = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

$$(87-3)$$

$$S_2 = \frac{1}{n_2-1}\sum_{j=1}^{n_2}(X_{2j}-\bar{X}_2)(X_{2j}-\bar{X}_2)' = \begin{bmatrix} S'_{11} & S'_{12} \\ S'_{21} & S'_{22} \end{bmatrix}$$

وبما إننا نفترض أن مصفوفتي التباين المشترك في المجموعتين G_1 و G_2 متساويتان، أي $(V_1 = V_2 = V)$، فإننا نقوم بتركيب مصفوفتي التباين المشترك S_1 و S_2 لنحصل منهما على مصفوفة واحدة، تصلح لأن تكون تقديراً غير متحيز للمصفوفة V المعرفة في العلاقة (3-71) .

وإن أفضل تركيب لهما هو المتوسط المثقل للمصفوفتين S_1 و S_2 بدرجتي الحرية لحجمي العينتين (n_1-1) و (n_2-1) والمسمى بالتباين المدمج $(Pooled)$ ويرمز له بـ S_p ويعرف بالعلاقة :

$$S_p = \left[\frac{(n_1-1)}{(n_1-1)+(n_2-1)}\right]S_1 + \left[\frac{(n_2-1)}{(n_1-1)+(n_2-1)}\right]S_2 \qquad (88-3)$$

$$S_p = \frac{(n_1-1)S_1 + (n_2-1)S_2}{\mathrm{n}_1 + n_2 - 2} \qquad (89-3)$$

وهو التقدير غير المتحيز للمصفوفة V، إذا كانت بيانات المصفوفتين X_1 و X_2 مأخوذة من عينتين عشوائيتين مسحوبتين من المجموعتين G_1 و G_2 على الترتيب .

وبذلك نكون قد حصلنا من بيانات هاتين العينتين على تقديرين غير متحيزين لـ $\boldsymbol{\mu_1}$ و $\boldsymbol{\mu_2}$ وللمصفوفة V، وبناء على ذلك يمكننا حساب القيم التقديرية لـ ℓ من العلاقة (3-78) كما يلي:

$$\ell = S_p^{-1}(\bar{X}_1 - \bar{X}_2) \qquad (90-3)$$

ومنها يمكننا إيجاد تقدير لتابع فيشر التمييزي الخطي من العلاقة:

$$\tilde{Y} = \tilde{\ell}'X = (\bar{X}_1 - \bar{X}_2)' * S_p^{-1} * X \qquad (91-3)$$

ومنها نحصل على تقديري متوسطي Y في العينتين من العلاقتين :

$$\tilde{y}_1 = \tilde{\ell}' * \bar{X}_1 \qquad \tilde{y}_2 = \tilde{\ell}' * \bar{X}_2 \qquad (92-3)$$

ومنهما نحصل على تقدير القيمة الوسطى لها من علاقة شبيهه بالعلاقة (3-81) وهي:

$$\tilde{m} = \frac{1}{2}(\tilde{y}_1 + \tilde{y}_2) = \frac{1}{2}\ell'(\bar{X}_1 + \bar{X}_2)' = \frac{1}{2}(\bar{X}_1 - \bar{X}_2)' * S_p^{-1} * (\bar{X}_1 + \bar{X}_2) \quad (93-3)$$

وهكذا نجد أن قاعدة التصنيف حسب التابع التمييزي (لفيشر) في العينات تصبح كما يلي:

لنأخذ عنصراً $X_0 = \begin{bmatrix} X_{01} \\ X_{02} \end{bmatrix}$ بشكل عشوائي من ذلك المجتمع، ثم نقوم بحساب تقدير قيمة التابع التمييزي y_0 من العلاقة (3-91):

$$\tilde{y}_0 = (\bar{X}_1 - \bar{X}_2)' * S_p^{-1} * X_0$$

ثم نقارن قيمة $\tilde{y}_0$ مع تقدير القيمة الوسطى $\tilde{m}$ ونتخذ قرار التصنيف كما يلي :

إذا كانت $\tilde{y}_0 \geq \tilde{m}$ نصنف X_0 في المجموعة G_1

وإذا كانت $\tilde{y}_0 < \tilde{m}$ نصنف X_0 في المجموعة G_2 $\quad (94-3)$

أو بعبارة أخرى:

إذا كانت $\tilde{y}_0 - \tilde{m} \geq 0$ نصنف X_0 في المجموعة G_1

وإذا كانت $\tilde{y}_0 - \tilde{m} < 0$ نصنف X_0 في المجموعة G_2 $\quad (95-3)$

ملاحظة: يمكننا إعادة صياغة علاقة معيار (فيشر) في (3-75) بعد تعويض d و V^{-1} بتقديرها $\tilde{d} = (\bar{x}_1 - \bar{x}_2)$ و $\tilde{V}^{-1} = S_p^{-1}$، فنحصل على ما يلي:

$$Max \frac{(\ell' d)^2}{\ell' * S_p * \ell} = \tilde{\boldsymbol{d}}' * \boldsymbol{S_p^{-1}} * \tilde{\boldsymbol{d}} = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' * \boldsymbol{S_p^{-1}} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) = D^2 \qquad (96-3)$$

ونعلم أن هذا يحدث عندما تكون $\tilde{\boldsymbol{d}} = \mathbf{C}\boldsymbol{S_p^{-1}}\boldsymbol{d} = \mathbf{C}\boldsymbol{S_p^{-1}}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$، وهذا يعني أن القيمة العظمى للنسبة (3-75) في حاله مجموعتين تساوي مربع المسافة بين مركزي العينتين المسحوبتين من هاتين المجموعتين D^2 .

وهذا يفيدنا في اختبار فيما إذا كان الفرق $(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$ معنوياً أم لا. لأنه يمكن النظر إلى اختبار الفرق بين أشعة المتوسطات كاختبار معنوية الحد الفاصل الذي حصلنا عليه .

لذلك نفترض أن القياسات في المجموعتين G_1 و G_2 خاضعة للتوزيع الطبيعي ولهما تباين مشترك موحد، V ونضع الفرضيتين (العدم H_0 والبديلة H_1) كما يلي:

$$H_0 : \boldsymbol{\mu_1} = \boldsymbol{\mu_2} \qquad H_1 : \boldsymbol{\mu_1} \neq \boldsymbol{\mu_2} \qquad (97-3)$$

ويتم اختبار H_0 باستخدام مؤشر الاختبار التالي:

$$F = \left[\frac{\mathrm{n}_1 + n_2 - P - 1}{(\mathrm{n}_1 + n_2 - 2) * P}\right]\left(\frac{\mathrm{n}_1 * n_2}{\mathrm{n}_1 + n_2}\right) * D^2 \qquad (98-3)$$

وهو يخضع لتوزيع $F(x)$ بدرجتي حرية $(v_1 = P)$ و $(v_2 = \mathrm{n}_1 + n_2 - P - 1)$، وعندما نرفض H_0 بمستوى دلالة $\propto$ يمكننا أن نستخلص أن الفصل بين المجموعتين G_1 و G_2 يعتبر معنوياً .

ملاحظة: إن الفصل المعنوي لا يؤدي بالضرورة إلى تصنيف جيد. لأن فعالية أسلوب التصنيف يمكن أن تكون مستقلة عن أي من اختبارات الفصل بين المجموعات . ولكن بالمقابل إذا كان الفصل غير معنوي فإن البحث عن تصنيف نافع يعتبر ضياعاً للوقت .

** **المعْيرة:** إن شعاع الأمثال الأولى $\ell(\ell_1, \ell_2)$ أو تقديره $\tilde{\ell}(\tilde{\ell}_1, \tilde{\ell}_2)$ الذي حصلنا عليه من إحدى العلاقتين (3−77) أو (3−87) ليس وحيداً، بل يمكن الحصول منه على لا نهاية من الحلول بضربه بأي عدد ثابت $C \neq 0$ فنحصل على الحلول $C\ell$ أو $C\tilde{\ell}$ التي تعظم قيمة النسبة (3−74).

وحتى نستطيع تفسير هذه الحلول من العلاقة (3−77) أو(3−78)، نقوم بإجراء معيرة للشعاع $\ell = \begin{bmatrix} \ell_1 \\ \ell_2 \end{bmatrix}$، بحيث نحصل على حل خاص يكون طوله مساوياً للواحد، ولإجراء هذه المعيرة يمكن اتباع إحدى الطريقتين التاليتين:

الطريقة الأولى: تتلخص بتقسيم عناصر الشعاع $\ell = \begin{bmatrix} \ell_1 \\ \ell_2 \end{bmatrix}$ على طوله $\|\ell\|$ فنحصل على شعاع واحدي e، يكون طوله مساوياً للواحد، ونحسب عناصره من العلاقة:

$$e = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} \dfrac{\ell_1}{\|\ell\|} \\ \dfrac{\ell_2}{\|\ell\|} \end{bmatrix} \qquad \mathbf{(99-3)}$$

ولذلك نقوم أولاً بحساب طول الشعاع ℓ من العلاقة :

$$\|\ell\| = \sqrt{\ell_1^2 + \ell_2^2} = \sqrt{\ell' * \ell} \qquad \mathbf{(100-3)}$$

ثم نقوم بحساب مركبات الشعاع الواحدي e من العلاقتين :

$$e_1 = \frac{\ell_1}{\|\ell\|} \qquad e_2 = \frac{\ell_2}{\|\ell\|} \qquad (101-3)$$

فنحصل على الشعاع الواحدي e المعرف في (3−99) والذي طوله يساوي الواحد لأن:

$$\|e\| = \sqrt{e' * e} = \sqrt{e_1^2 + e_2^2} = \sqrt{\left(\frac{\ell_1}{\|\ell\|}\right)^2 + \left(\frac{\ell_2}{\|\ell\|}\right)^2} = 1$$

الطريقة الثانية: وهي تعتمد على جعل تباين تابع (فيشر) مساوياً للواحد، أي جعل العلاقة (3−73) تساوي الواحد:

$$vor(y) = \sigma_y^2 = \ell' * V * \ell = 1 \qquad (102-3)$$

واعتماداً على خواص المصفوفات المتناظرة والمحددة إيجابياً نكتب المصفوفة المتناظرة والمحددة إيجابياً V، كما يلي: $\left(V = V^{\frac{1}{2}}V^{\frac{1}{2}}\right)$ ونعالج العلاقة (3−102) كما يلي:

$$\sigma_y^2 = \ell' * V * \ell = \ell' * V^{\frac{1}{2}} * V^{\frac{1}{2}} * \ell = \left(V^{\frac{1}{2}} * \ell\right)' * \left(V^{\frac{1}{2}} * \ell\right) = 1 \qquad (103-3)$$

وإذا رمزنا للشعاع الذي ضمن القوس الثاني $\left(V^{\frac{1}{2}} * \ell\right)$ بالرمز e يكون لدينا:

$$e = V^{\frac{1}{2}} * \ell \qquad \text{و} \qquad e' = \left(V^{\frac{1}{2}} * \ell\right)' \qquad (104-3)$$

ومنه نحصل على أن:

$$\sigma_y^2 = e' * e = \|e\|^2 = 1 \qquad (105-3)$$

أي أن الشعاع الجديد $e = V^{\frac{1}{2}} * \ell$ المعرف بالعلاقة (3‒104) هو شعاع واحدي (طوله واحد)، وبذلك نحصل على حل خاص ممعير e من كل حل عادي ℓ، ومن المصفوفة $V^{\frac{1}{2}}$ واعتماداً على العلاقة (2‒58) في الملحق (2) نجد ما يلي :

$$e = V^{\frac{1}{2}} * \ell = E * \Lambda^{\frac{1}{2}} * E' * \ell \qquad (106-3)$$

ومنها يمكن استخلاص ℓ بدلالة e كما يلي:

$$\ell = V^{-\frac{1}{2}} * e = E * \Lambda^{-\frac{1}{2}} * E' * e \qquad (107-3)$$

حيث أن E هي مصفوفة الأشعة الذاتية الممعيرة والمتعامدة للمصفوفة V. وأن $\Lambda^{\frac{1}{2}}$ هي المصفوفة القطرية لجذور القيم الذاتية للمصفوفة V.

أي أنه إذا كان لدينا حل ممعير $e = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$ لتابع (فيشر) المعياري، فإنه يمكننا منه استخلاص الحل الخاص الخام ℓ لذلك التابع من العلاقة :

$$\ell = V^{-\frac{1}{2}} * e \qquad (107a-3)$$

ملاحظة: ينصح بإجراء المعيرة على ℓ إذا كان المتحولان X_1 و X_2 معياريين .

مثال (3‒2): في دراسة ميدانية قام بها *[Bouma, B. N.+ others 1975]* لتقدير معدل انتشار مرض الناعور A عند النساء في الولايات المتحدة [المصدر(3) ص477] سحب عينتين من النساء كمايلي:

العينة الأولى: تتألف من $n_1 = 30$ امرأة، وتم سحبها من مجموعة النساء العاديات اللاتي لا يحملن جينات مرض الناعور A ونرمز لها بـ G_1 .

العينة الثانية: تتألف من $n_2 = 22$ امرأة، وتم سحبها من مجموعة النساء اللاتي يحملن مرض الناعور A (بنات الحاملات، أمهات أبناء أو بنات مرضى) وسميت هذه المجموعة بمجموعة الحوامل الإجبارية ونرمز لها بـ G_2 .

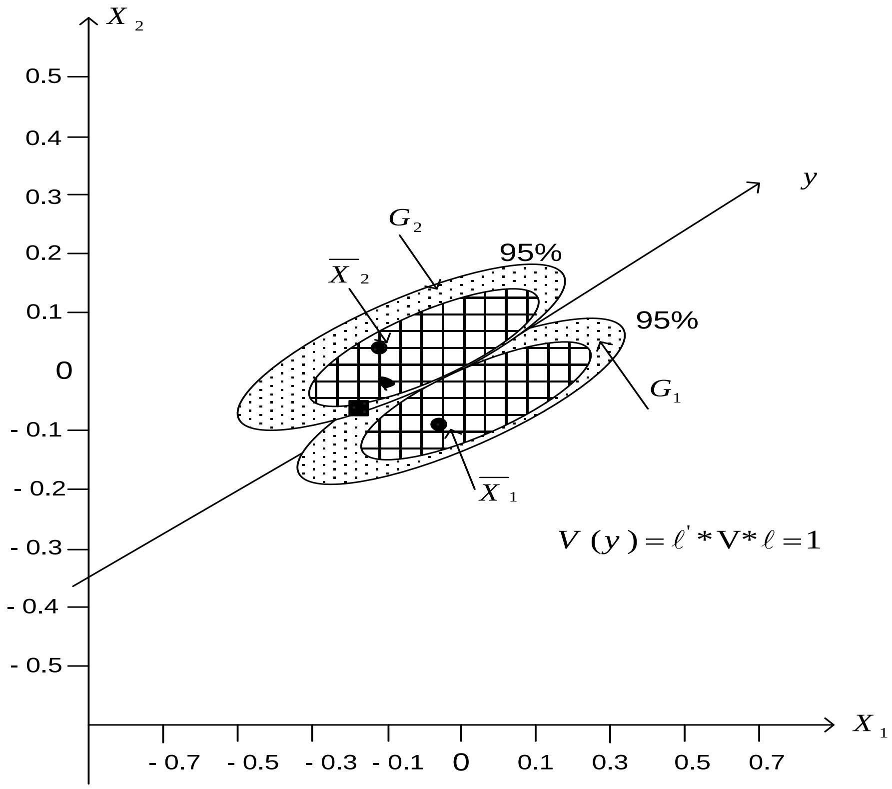
وللتمييز بين هاتين المجموعتين تم الاعتماد على متحولين X_1 و X_2 تم تعريفها وقياسها وتحويلها من نتائج تحليل الدم كما يلي:

$$X_1 = log_{10}(AHF\ Ativity) = \text{فعالية عامل الالتهاب}$$

$$X_2 = log_{10}(AHF - like\ antigen) = \text{المستضد : فعالية المقاومة}$$

ومن أزواج هذين المتحولين (X_1, X_2) في كل مجموعة حصل على القياسات المقابلة لذلك، والتي نعرضها ضمن قطعين ناقصين على الشكل (3‒7) ولكل مجموعة على حدة، القطع الصغير يضم 50% من المشاهدات، والكبير يضم 95% من المشاهدات . وبعد اختبار خضوع هذه المشاهدات للتوزيع الطبيعي المتعدد في كل مجموعة تم حساب متوسطيهما فكانا كما يلي:

$$\bar{X}_1 = \begin{bmatrix} -0.0065 \\ -0.0390 \end{bmatrix} \qquad \bar{X}_2 = \begin{bmatrix} -0.2483 \\ 0.0262 \end{bmatrix}$$



الشكل (3–7) توزيع قياسات X_1 و X_2

وبعد حساب مصفوفتي التباين المشترك لهما S_1 و S_2 في المجموعتين G_1 و G_2 تم حساب المصفوفة المدمجة للتباين المشترك S_p من العلاقة (3–86)، ثم تم حساب مقلوبها فكان يساوي كما يلي:

$$S_p^{-1} = \begin{bmatrix} 131.158 & -90.423 \\ -90.423 & 108.147 \end{bmatrix}$$

وبعدها نقوم بحساب تابع فيشر التمييزي من العلاقة (3–88) التالية:

$$\tilde{y} = \tilde{\ell}' * X = (\bar{X}_1 - \bar{X}_2)' S_p^{-1} * X =$$

لذلك نحسب أولاً الشعاع $[\bar{X}_1 - \bar{X}_2]$ فنجد أنه يساوي:

$$[\bar{X}_1 - \bar{X}_2] = \begin{bmatrix} (\bar{X}_{11} - \bar{X}_{21}) \\ (\bar{X}_{12} - \bar{X}_{22}) \end{bmatrix} = \begin{bmatrix} (-0.0065 + 0.2483) \\ (-0.0390 - 0.0262) \end{bmatrix} = \begin{bmatrix} 0.2418 \\ -0.0652 \end{bmatrix}$$

ثم نعوض ذلك في المعادلة السابقة فنجد أن:

$$\tilde{y} = [0.2148 \quad -0.0652] \begin{bmatrix} 131.158 & -90.423 \\ -90.423 & 108.147 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

$$\tilde{y} = [37.61 \quad -28.921] \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

$$\bar{y} = 37.61 X_1 - 28.921 X_2 \quad \Longrightarrow \quad \tilde{\ell} = \begin{bmatrix} 37.61 \\ -28.921 \end{bmatrix}$$

وهو تابع فيشر التمييزي لهذه المسألة، وهو المستقيم المرسوم على الشكل (3–6) السابق. ثم نقوم بحساب تقديري المتوسطين μ_{1y} و μ_{2y} فنجد أن:

$$\tilde{\mu}_{1y} = \tilde{\ell}' * \bar{X}_1 = [37.81 \quad -28.921] \begin{bmatrix} -0.0065 \\ -0.0395 \end{bmatrix} = 0.88$$

$$\tilde{\mu}_{2y} = \tilde{\ell}' * \bar{X}_2 = [37.81 \quad -28.921] \begin{bmatrix} -0.2483 \\ 0.0262 \end{bmatrix} = -10.10$$

ثم نحسب القيمة الوسطى لهما فنجد أن:

$$\tilde{m} = \frac{1}{2}\left(\tilde{\mu}_{1y} + \tilde{\mu}_{2y}\right) = \frac{1}{2}(0.88 - 10.10) = -4.61$$

ولاستخدام هذه النتائج في تصنيف النساء على المجموعتين G_1 (العاديات) و G_2 (الحاملات لمرض الناعور A) نفترض أن نتائج تحليل الدم لإحدى النساء كانت كما يلي: $X_0 = \begin{bmatrix} -0.210 \\ -0.044 \end{bmatrix}$ أي أن:

$$x_{01} = -0.210 \qquad\qquad x_{02} = -0.044$$

نعوض ذلك في تابع فيشر التمييزي فنجد أن قيمته تساوي:

$$\tilde{y}_0 = \tilde{\ell}' x_0 = [37.61 \quad -28.921]\begin{bmatrix} -0.210 \\ -0.044 \end{bmatrix} = -6.62$$

وعند مقارنة قيمة $\tilde{y}_0$ مع $\overline{m}$ نجد أن:

$$\tilde{y}_0 < \tilde{m} \qquad \Leftrightarrow \qquad -6.62 < -4.61$$

لذلك نصنف هذه المرأة ضمن المجموعة G_2 أي ضمن مجموعة حاملات مرضى الناعور A. ولقد أشرنا إلى هذه الحالة على الشكل (3−7) بإشارة نجمة. ومن الشكل نلاحظ أنها تقع ضمن قطع الـ 50% لمجموعة الحاملات G_2 وضمن قطع الـ95% لمجموعة النساء العاديات G_1، وهذا يعني أن هذا التصنيف ليس واضح المعالم .

والآن نقوم بمعيرة شعاع الثوابت ℓ في تابع فيشر التمييزي وذلك باستخدام العلاقة:

$$\tilde{e}^* = \frac{\tilde{\ell}}{\sqrt{\tilde{\ell}'\tilde{\ell}}} = \frac{\tilde{\ell}}{\|\tilde{\ell}\|}$$

لذلك نحسب طول الشعاع $\tilde{\ell}$ من العلاقة فنجد أن:

$$\|\ell\| = \sqrt{\ell' * \ell} = \sqrt{\tilde{\ell}_1^2 + \tilde{\ell}_2^2} = \sqrt{(37.61)^2 + (-28.921)^2} = \sqrt{2250.936} = 47.444$$

ثم نقسم مركبات $\tilde{\ell}$ على $\|\ell\|$ فنحصل على الشعاع الممعير التالي:

$$\tilde{\text{e}} = \begin{bmatrix} \dfrac{\tilde{\ell}_1}{\|\tilde{\ell}\|} \\ \dfrac{\tilde{\ell}_2}{\|\tilde{\ell}\|} \end{bmatrix} = \begin{bmatrix} \dfrac{37.61}{47.444} \\ \dfrac{-28.921}{47.444} \end{bmatrix} = \begin{bmatrix} 0.7927 \\ -0.6096 \end{bmatrix} \approx \begin{bmatrix} 0.79 \\ -0.61 \end{bmatrix}$$

وبذلك تصبح المعادلة الممعيرة لتابع فيشر التمييزي كما يلي:

$$\tilde{y}^* = \tilde{e}' * X = 0.791X_1 - 0.61X_2$$

ومنه نستنتج أن الأهمية النسبية لـ X_1 أكبر من X_2 في التأثير على $\tilde{y}^*$، وبغض النظر عن الإشارة فإن الفرق بينهما صغيراً .

وحتى نطبق عملية التصنيف بعد هذه المعيرة، نقوم بحساب نقطة القيمة الوسطى الممعيرة من العلاقة:

$$\tilde{m}^* = \frac{1}{2}\left(\mu_{1y}^* + \mu_{2y}^*\right) = \frac{1}{2}(\tilde{e}'\bar{X}_1 + \tilde{e}'\bar{X}_2) =$$

$$\tilde{m}^* = \frac{1}{2}\left\{\left[(0.79, -0.61)\begin{bmatrix} -0.0065 \\ -0.0390 \end{bmatrix} + (0.79, -0.61)\begin{bmatrix} -0.2483 \\ 0.0262 \end{bmatrix}\right]\right\} = -0.0973$$

وإذا كنا نريد أن نصنف المشاهدة السابقة $X_0' = [-0.210\ ,\ \ -0.044]$ نعوضها في تابع فيشر التمييزي الممعير فنحصل على أن قيمته تساوي:

$$\tilde{y}^* = e^{*\prime} * X = (0.791\ , -0.61)\begin{bmatrix}-0.210\\-0.044\end{bmatrix} = -0.1397$$

ثم نقارن هذه القيمة مع القيمة الوسطى الممعيرة$\widetilde{m}^*$ فنجد أن:$\tilde{y}^* < \widetilde{m}^*$ (لأن$(-0.1397 < -0.097)$)

لذلك نصنف هذه المشاهدة ضمن المجموعة G_2 (مجموعة الحاملات) .

ملاحظة: يمكن حساب قيمتي $\tilde{y}_0^*$ و$\widetilde{m}_0^*$ الممعيرتين مباشرة من القيمتين الأصليين لهما $\tilde{y}_0$ وm_0، وذلك بضربهما بثابت المعيرة، وهو مقلوب طول الشعاع ℓ والذي يساوي:

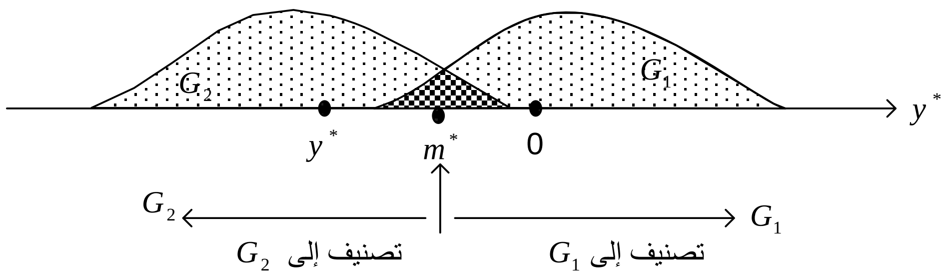
فنجد أن:

$$C = \frac{1}{||\ell||} = \frac{1}{47.44} = 00211$$

$$\tilde{y}^* = C * \tilde{y} = 0.0211(-6.62) = -0.1397$$

$$\widetilde{m}^* = C * \widetilde{m} = 0.0211(-4.61) = -0.0973$$

ويمكن توضيح عملية التصنيف بيانياً كما يلي:



الشكل (3−8) محور التصنيف y

- **إضافات أخرى للبرهان على تابع (فيشر)** :هناك أساليب وطرائق أخرى لتحديد قيمة الشعاع ℓ في العلاقة (3−74) التي تجعل النسبة F أكبر ما يمكن :
- **الطريقة الأولى:** لتطبيق هذه الطريقة نقوم بكتابة المعيار F كما يلي:

$$F = \frac{\left(\ell' d\right)^2}{\ell' * V * \ell} = \left(\boldsymbol{\ell}'\boldsymbol{d}\right)^2 * \frac{1}{\ell' * V * \ell} \qquad (108-3)$$

ثم نقوم باشتقاق F بالنسبة لـ ℓ (حسب قواعد اشتقاق المصفوفات)، ونضع ذلك المشتق مساوياً للصفر فنجد أن:

$$\frac{dF}{d\ell} = 2\left(\ell' d\right) * \boldsymbol{d} * \frac{1}{\ell' * V * \ell} - \frac{2V * \ell}{\left(\ell' * V * \ell\right)^2} * \left(\ell' d\right)^2 = 0$$

نختصر على (2) ثم نخرج المقدار العددي$\left(\frac{\ell' d}{\ell' * V * \ell}\right)$ خارج قوس فنحصل على أن:

$$\frac{\ell' d}{\ell' * V * \ell}\left[d - \left(\frac{\ell' d}{\ell' * V * \ell}\right) * V * \ell\right] = 0 \qquad (109-3)$$

وبما أن المقدار$\left(\frac{\ell' d}{\ell' * V * \ell}\right)$ هو عبارة عن عدد حقيقي سلمي لذلك نرمز له بـ K فنحصل على أن:

$$K[d - K * V * \ell] = 0 \qquad (110-3)$$

وبعد الاختصار على K نحصل على المعادلة التالية:

$$d = K * V * \ell \qquad (111-3)$$

ولتحديد قيمة ℓ نضرب الطرفين من اليسار بـ V^{-1} ثم نقسم على k فنحصل على أن ℓ يساوي:

$$\ell = \frac{1}{K} V^{-1} * d = C * V^{-1} * \boldsymbol{d} \qquad : (c = \frac{1}{k}) \qquad (112-3)$$

وبما أن $\boldsymbol{d} = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$ فإننا نستنتج أن:

$$\ell = C * V^{-1} * (\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) \qquad (113-3)$$

وهي نفس العلاقة التي حصلنا عليها في (3−77)، ومنها نستنتج أن الشعاع ℓ يتناسب طرداً مع الشعاع $(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) * V^{-1}$، وهنا يعطينا لا نهاية من الحلول المتناسبة لـ ℓ وذلك حسب قيم $C \neq 0$.

وإذا قمنا بوضع $C = 1$ نحصل على الحل الخاص التالي:

$$\ell = V^{-1} * (\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) \quad \Longrightarrow \quad \ell' = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * V^{-1} \qquad (114-3)$$

وبتعويض ذلك في تابع (فيشر) نحصل على العلاقة التالية:

$$y = \ell' * X = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})' * V^{-1} * X \qquad (115-3)$$

وفي حالة العينات نقدرها بالعلاقة التالية:

$$\tilde{y} = \tilde{\ell}' * X = (\bar{X}_1 - \bar{X}_2) * S_p^{-1} * X \qquad (116-3)$$

ولاستخدام هذه العلاقة في عمليات تصنيف المشاهدات على المجموعتين نقوم بحساب القيمة الوسطى للمتوسطين μ_{1y} و μ_{2y} أو لتقديرهما $\bar{X}_{1y}$ و $\bar{X}_{2y}$ فنحصل على :

$$\widetilde{m} = \frac{1}{2}\left(\tilde{\mu}_{1y} + \tilde{\mu}_{2y}\right) = \frac{1}{2}\left(\bar{X}_{1y} + \bar{X}_{2y}\right) \qquad (117-3)$$

وبعدها نطبق القاعدة (3−94) على أية مشاهدة مثل X_0 .

- **الطريقة الثانية:** ومن جهة أخرى نلاحظ أنه إذا كنا نشترط أن يكون $\sigma^2_{(y)} = 1$ فإن ذلك يعني أن نضيف إلى المعيار F الشرط التالي:

$$\sigma^2_{(y)} = \ell' * V * \ell = 1 \qquad (118-3)$$

وعندها فإن معيار (فيشر) في (3−74) يأخذ شكلاً مبسطاً كما يلي:

$$F = \frac{(\ell' * d)^2}{1} \Longrightarrow Max \qquad \boldsymbol{(119-3)}$$

ويصبح المطلوب منا أن نحسب عناصر الشعاع $\ell = \begin{bmatrix} \ell_1 \\ \ell_2 \end{bmatrix}$ التي تجعل المعيار F في (3−119) أكبر ما يمكن ضمن الشرط الجديد (3−118) .

ولإيجاد الشعاع ℓ نشكل تابع (لاغرانج) من التابع (3− 119) والشرط (3−118) كما يلي:

$$L = (\ell' * \boldsymbol{d})^2 - \lambda(\boldsymbol{\ell}' * V * \boldsymbol{\ell} = 1) \qquad (120-3)$$

ثم نشتق التابع L بالنسبة لـ ℓ (حسب قواعد اشتقاق المصفوفات) ونضع ذلك المشتق مساوياً للصفر فنحصل على أن:

$$\frac{\partial L}{\partial \ell} = 2(\boldsymbol{\ell}' * \boldsymbol{d}) * \boldsymbol{d} - \lambda * 2\boldsymbol{V} * \boldsymbol{\ell} = 0 \qquad (121-3)$$

وبعد الاختصار على (2) نحصل منها على المعادلة التالية:

$$(\boldsymbol{\ell}' * \boldsymbol{d}) * d - \lambda * \boldsymbol{V} * \boldsymbol{\ell} = 0 \qquad (122-3)$$

وبما أن $\ell' * d$ هو عدد حقيقي سلمي لأن :

$$\boldsymbol{\ell}' * \boldsymbol{d} = (\ell_1 , \ell_2) \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \ell_1 d_1 + \ell_2 d_2 = C_1 = \text{عدد سلمي}$$

ولذلك نرمز له بـ C_1 فنجد أن العلاقة تأخذ الشكل التالي:

$$C_1 * d = \lambda * V * \ell \qquad (123-3)$$

ثم نضرب الطرفين من اليسار بـ V^{-1} ونقسمهما على λ فنحصل على أن الشعاع ℓ يساوي:

$$\ell = \frac{C_1}{\lambda} V^{-1} * d = CV^{-1} * \boldsymbol{d} \qquad : \left(C = \frac{C_1}{\lambda} \right) \qquad (124-3)$$

وبما أن $d = (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$ فإن ℓ يساوي:

$$\boldsymbol{\ell} = CV^{-1}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) \qquad (125-3)$$

وهي نفس العلاقة التي حصلنا عليها في (3-77)، أي أن ℓ يتناسب طرداً مع الشعاع، $V^{-1} * (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$ وأن العلاقة (3-125) تعطينا لا نهاية من الحلول. ولكن بوضع $C = 1$ نحصل على حل خاص له هو:

$$\boldsymbol{\ell} = V^{-1}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) \qquad (126-3)$$

وفي حالة العينات نحسب تقدير ℓ من العلاقة:

$$\tilde{\boldsymbol{\ell}} = S_p^{-1}(\bar{X}_1 - \bar{X}_2) \qquad (127-3)$$

وبتعويض ذلك في تابع (فيشر) نحصل على تابع (فيشر) التمييزي التالي:

$$y = \tilde{\boldsymbol{\ell}}' X = (X_1 - X_2)' * S_p^{-1} * X \qquad (128-3)$$

ثم نتعامل معه كالعادة ونحسب القيمة الوسطى لـ y، ثم نطبق القاعدة من أجل تصنيف أي عنصر جديد من المجتمع مثل X_0 .