الفصل الثاني

التحليل التمييزي البسيط (حسب متحول واحد ومجموعتين)

2-1 تمهيد:

يعتبر التحليل التمييزي البسيط أساساً لأنواع التمييز الأخرى، لأنه يمكن أن يتم تحويل معظم الأنواع الأخرى إلى التحليل التمييزي البسيط . ولاستعراض هذا النوع البسيط نأخذ مجتمع القروض المصرفية الذي يقدمها أحد المصارف إلى المواطنين ونعرضها كما يلي:

لنفرض أن مدير المصرف لاحظ أن الكثير من القروض الممنوحة متعثرة السداد (أكثر من ستة أشهر)، فطلب من المسؤولين عنها تحديد أسباب التعثر لمعالجة هذه الظاهرة فقالوا له: توجد أسباب كثيرة لذلك ومنها (الدخل والعمر والملاءة والجنس والمهنة والخبرة والمنافسة والالتزام والأخلاق ...الخ)، ولكن المدير طلب منهم تحديد مؤشر واحد أو مؤشرين لاستخدامهما في قرار الموافقة على منح القروض، فاستقر أمرهم الأول على اعتبار (الدخل الشهري للمقترض X) هو أهم الأسباب المؤدية إلى تعثر القروض، لأنه إذا كان دخل المقترض ضعيفاً، فإن قدرته على تسديد الأقساط تكون ضعيفة، وبالتالي فإن قرضه يكون عرضة للتعثر .

وللتأكد من ذلك قام فريق العمل بسحب عينة عشوائية من القروض المتعثرة بحجم $n_1 = 8$، وسحب عينة عشوائية من القروض غير المتعثرة بحجم $n_2 = 7$، ثم قام بتصنيف عناصر هاتين العينتين حسب دخل المقترض في كل منها ثم معالجتها، فكانت بيانات ومتوسطات وتباينات دخول أصحاب هذه القروض كما يلي:

جدول (2-1): بيانات عينيتي القروض المتعثرة وغير المتعثرة . (من *Lesson2: DA SPSS* بتصرف)

	مجموعة القروض المتعثرة G_1								مجموعة القروض غير المتعثرة G_2						
	عناصر عينة القروض المتعثرة : $n_1 = 8$								عناصر عينة القروض غير المتعثرة : $n_2 = 7$						
الرقم	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7
دخل صاحب القرض	56	56	45	56	29	56	56	30	72	72	56	72	64	64	56
المتوسط والتباين	$\bar{x}_1 = 48$ $\quad S_1^2 = (12.036)^2$								$\bar{x}_2 = 65.14$ $\quad S_2^2 = (7.192)^2$						

وكان المتوسط العام المثقل للدخول في هاتين العينتين (المتعثرين وغير المتعثرين) يساوي :

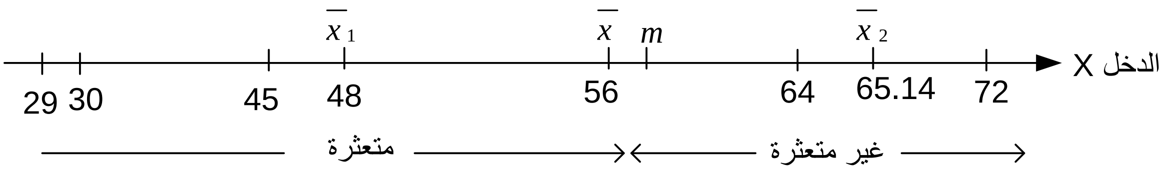
$$\bar{x} = \frac{8(48) + 7(65.14)}{15} = 56$$

وإن التباين العام والانحراف المعياري يساويان :

$$S^2 = (12.153)^2 \quad , \quad S = 12.153$$

ويمكن حساب القيمة الوسطى للمتوسطين $\bar{x}_1$ و $\bar{x}_2$ من العلاقة : $m = \frac{\bar{x}_1+\bar{x}_2}{2} = 56.57$

إن هذه المتوسطات أو المراكز ترسم على المحور المخصص للدخل X على الشكل التالي:



الشكل رقم (2-1): محور توزيع القروض حسب الدخل

وهنا تساؤل المسؤولون: كيف سنفرز القروض حسب الدخل فقط إلى قروض متعثرة وقروض غيرمتعثرة ؟ وماهو المعيار الذي يمكن اعتماده لتحقيق ذلك الفرز بأقل خطأ ممكن ؟

إن الجواب السهل والبسيط على ذلك هو أن نستفيد من المتوسط العام $\bar{x}$، ونعتبره الحد الفاصل بين هاتين المجموعتين ونقول:

إذا كان $X \leq \bar{x}$ فإن القرض يكون متعثراً

أما إذا كان $X > \bar{x}$ فإن القرض يكون غير متعثر .

وفي مثالنا هذا نلاحظ أن $\bar{x} = 56$، وهذا يعني أن جميع القروض المتعثرة $(n = 8)$ ستصنف على أنها متعثرة وهذا تصنيف صحيح .

ولكننا سنقع في خطأ ملحوظ عند تصنيف القروض غير المتعثرة، حيث سنعتبر القرضين (3) و (7) من القروض غير المتعثرة قروضاً متعثرة (وهذا تصنيف خاطئ). وهكذا نجد إن هذا المعيار (اعتبار المتوسط العام $\bar{x}$ هو الحد الفاصل) هو معيار معقول ولكنه يحتمل الخطأ، فعند دراسة الطلبات الجديدة للزبائن فإنه حسب هذا المعيار سنعتبر كل من كان دخله $X \leq 56$ من أصحاب القروض المتعثرة، وهذا أمر مشكوك فيه لأن بيانات العينة الثانية تفيدنا أن نسبة لا بأس بها من أصحاب الدخول $X \leq 56$ هم من أصحاب القروض غير المتعثرة (وليس المتعثرة) .

إن الوقوع في هذا الخطأ يحدث لعدة أسباب أهمها:

1- إذا كانت الدراسة لا تستند إلى خصائص المجتمع ولا تعتمد على التوزيعات التكرارية الاحتمالية للدخول في كل مجموعة في المجتمع .

2- إذا كانت العينات المسحوبة غير عشوائية وكان حجمها الكلي صغيراً $(n < 30)$.

3- إذا كانت حجوم العينات الطبيعية لا تتناسب مع حجوم المجموعات . أي إذا كانت نسبة حجم كل منها لا يتوافق مع نسبة حجم المجموعة في المجتمع .

فمثلاً: إذا كانت نسبة القروض المتعثرة 30% أي $(P_1 = 0.30)$ من القروض الكلية .

وكانت نسبة القروض غير المتعثرة 70% أي $(P_2 = 0.70)$ من القروض الكلية .

فإن نسبة حجمي العينتين يجب أن يتوافق مع هاتين النسبتين. أي يجب أن يكون :

$$\frac{n_1}{n} = P_1 \qquad \frac{n_2}{n} = P_2 \qquad (1-2)$$

4- إذا كانت حجوم العينات لا تأخذ بعين للاعتبار الانحرافات المعيارية داخل كل مجموعة، لأن حجوم العينات يجب أن تتناسب طرداً مع الانحرافات المعيارية الداخلية لتلك المجموعات أي يجب أن يكون :

$$n_1 = \frac{N_1}{N}\sqrt{S_{11}} \qquad n_1 = \frac{N_2}{N}\sqrt{S_{22}} \qquad (2-2)$$

حيث أن N_1 و N_2 هما عدد العناصر في المجموعتين وأن $(N_1 + N_2)$ = حجم المجتمع N.

5- إذا كانت حجوم العينات واحتمالات الخطأ لا تأخذ بعين الاعتبار تكلفة التصنيفات الخاطئة Misclassification cost . لأن حجوم العينات تتناسب عكساً مع تكلفة التصنيفات الخاطئة .

ولهذا يجب اتباع طرائق رياضية دقيقة لتحديد الحدود الفاصلة بين المجموعات، وأهم هذه الطرائق هي:

- طريقة التكلفة المتوقعة للتصنيفات الخاطئة ECM .
- طريقة بايز (Bayse) أو الاحتمالات الشرطية اللاحقة .
- طريقة الجوار الأقرب أو مسافة Mahalanobis .
- طريقة معيار فيشر Fishir's certion .
- طريقة تحليل التباين الأحادي أو المتعدد ANOVA أو MANOVA .

وسنتعرض لهذه الطرائق باختصار يتناسب مع التحليل التمييزي البسيط كما يلي :

2-2 طريقة التكلفة المتوقعة للتصنيفات الخاطئة ECM: (Expected Cost of Misclassification)

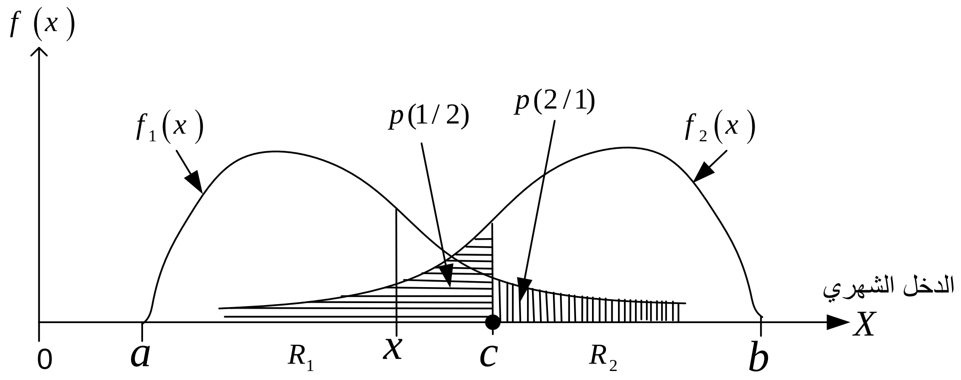
لنتابع مثالنا السابق حول القروض، ولنرمز للمجموعتين كما يلي:[Johnston A. P. 485 بتصرف وإضافة]

نرمز لمجموعة القروض المتعثرة بالرمز G_1 ولنسبتها في المجتمع P_1 .

نرمز لمجموعة القروض غير المتعثرة بالرمز G_2 ولنسبتها في المجتمع P_2 .

إن P_1 و P_2 يسميان بالاحتمالات السابقة . لأنهما يعطيانا احتمال انتماء أي عنصر جديد إلى إحدى هاتين المجموعتين ويحققان العلاقة $P_1 + P_2 = 1$.

ولنفترض الآن أننا جمعنا البيانات اللازمة عن جميع دخول المقترضين X في كل مجموعة على حدة، وحسبنا تكراراتها النسبية واستخلصنا منها التوزيع الاحتمالي للدخول في G_1 ورمزنا له بـ $f_1(x)$. والتوزيع الاحتمالي للدخول في G_2 ورمزنا له بـ $f_2(x)$. والمعرفين على مجال مشترك لهما هو $[a\,,b]$ وكان لهما الشكل التالي :



الشكل (2−2): التوزيعات التكرارية في G_1 و G_2

ونعلم أنه من خواص هذين التوزيعين إنهما يحققان الخاصتين التاليتين

$$f_1(x) \geq 0 \qquad \int_a^b f_1(x)dx = 1 \qquad (\mathbf{3-2})$$

$$f_2(x) \geq 0 \qquad \int_a^b f_2(x)dx = 1 \qquad (\mathbf{4-2})$$

والمطلوب تحديد النقطة الفاصلة C، التي تفرز القروض بأقل خطأ ممكن إلى مجموعتين: مجموعة المتعثرة G_1 على اليسار، ومجموعة غير المتعثرة G_2 على اليمين، والتي تقسم المجال $[a,b]$ إلى مجالين $[a,c]$ و $]c\ ,d]$.

بحيث يكون المجال $R_1 =]a\ ,\ c]$ مقابلاً للمجموعة المتعثرة G_1 .

ويكون المجال $R_2 =]c\ ,b]$ مقابلاً للمجموعة غير المتعثرة G_2 .

ولتحديد انتماء كل قرض إلى إحدى هاتين المجموعتين نقوم بحساب الاحتمالات الشرطية المقابلة لهما، فنجد أن احتمال أن يتم تصنيف القرض حسب الدخل لـ X خطأ في المجموعة الثانية G_2 وهو أصلاً من المجموعة الأولى G_1 يساوي:

$$P(2/1) = P[x \in R_2/G_1] = \int_{R_2} f_1(x)dx \qquad (5-2)$$

وهو يمثل بالمساحة التي يشكلها $f_1(x)$ فوق المجال R_2 وهي المساحة المظللة بخطوط عمودية . وهو يعبر عن احتمال أن ينتمي عنصر المجموعة G_1 خطأ إلى المجموعة G_2 ، أي هو احتمال أن يتم تصنيف عنصر المجموعة المتعثرة G_1 خطأ في المجموعة غير المتعثرة G_2 .

أما احتمال أن يتم تصنيف عنصر المجموعة G_2 خطأ في المجموعة G_1 فهو يساوي :

$$P(1/2) = P[x \in R_1/G_2] = \int_{R_1} f_2(x)dx \qquad (6-2)$$

وهو يمثل بالمساحة المظللة بخطوط أفقية والتي يشكلها التوزيع $f_2(x)$ فوق المجال R_1 .

ويطلق على الاحتمالين الشرطيين $P(1/2)$ و $P(1/2)$ وما شابههما مصطلح الاحتمالات اللاحقة، وهي تختلف عن الاحتمالات السابقة P_2 و P_1 التي ذكرناها سابقاً .

وبناء على ذلك يمكننا حساب احتمالات التصنيف الصحيح والتصنيف الخاطئ إلى كل مجموعة كمايلي:

إن احتمال أن يكون تصنيف العنصر x إلى المجموعة G_1 صحيحاً يساوي الجداء التالي =

= (احتمال أن يكون أصله من G_1)*(احتمال تصنيفه في G_1) .

ولنرمز لذلك الاحتمال بالرمز $P(1,1)$ فيكون لدينا :

$$P(1,1) = P[x \in R_1/G_1] * P(G_1) = P(1/1) * P_1 \qquad (7-2)$$

وإن احتمال أن يكون تصنيف العنصر x إلى المجموعة G_1 خاطئاً يساوي الجداء التالي =

(احتمال أن يكون أصله من G_2)*(احتمال تصنيفه خطأ في G_1) .

ولنرمز لذلك الاحتمال بالرمز $P(1,2)$ ويكون لدينا :

$$P(1,2) = P[x \in R_1/G_2] * P(G_2) = P(1/2) * P_2 \qquad (8-2)$$

أما احتمال أن يكون تصنيف العنصر x إلى المجموعة G_2 صحيحاً فيساوي الجداء التالي =

(احتمال أن يكون أصله من G_2)*(احتمال تصنيفه في G_2) .

ولنرمز لذلك الاحتمال بالرمز $P(2,2)$ ويكون لدينا :

$$P(2,2) = P[x \in R_2/G_2] * P(G_2) = P(2/2) * P_2 \qquad (9-2)$$

وإن احتمال أن يكون تصنيف العنصر x إلى المجموعة G_2 خطأً يساوي الجداء التالي =

(احتمال أن يكون أصله من G_1)*(احتمال تصنيفه خطأ في G_2) .

ولنرمز لذلك الاحتمال بالرمز $P(2,1)$ ويكون لدينا :

$$P(2,1) = P[x \in R_2/G_1] * P(G_1) = P(2/1) * P_1 \qquad (10-2)$$

ويمكن وضع هذه الاحتمالات الشرطية في جدول رباعي كما يلي:

جدول (2-2): الاحتمالات المتقاطعة للتصنيف في المجموعتين G_1 و G_2 :

G_2	G_1	المجموعات	الاحتمالات السابقة
$P(2/1) * P_1$	$P(1/1) * P_1$	G_1	P_1
$P(2/2) * P_2$	$P(1/2) * P_2$	G_2	P_2

ويصبح هدفنا الآن أن نجعل مجموع الاحتمالين المقابلين للحالتين الخاطئتين أصغر ما يمكن، وهو الذي يساوي:

$$P(2/1) * P_1 + P(1/2) * P_2 = TPM \qquad (11-2)$$

ويسمى هذا الاحتمال بالاحتمال الاجمالي للتصنيف الخاطئ، ويرمز له بـ TPM من الكلمات (Total Probability Misclassification)، وبناءً على العلاقتين (2-5) و(2-6) يمكن كتابته على الشكل التالي:

$$TPM = P_1 * \int_{R_2} f_1(x)dx + P_2 * \int_{R_1} f_2(x)dx \qquad (12\,a-2)$$

وعلينا الآن أن نجعل هذا الاحتمال الاجمالي TPM أصغر ما يمكن .

ولكن ذلك الأمر يتعلق بالتكلفة المترتبة على التصنيفات الخاطئة في كلا المجموعتين G_1 و G_2 .
فإذا افترضنا أن تكلفة التصنيف الصحيح في كل مجموعة تساوي الصفر. وإن تكلفة التصنيف الخاطئ في المجموعة الأولى تساوي $C(1/2)$. وإن تكلفة التصنيف الخاطئ في المجموعة الثانية تساوي $C(2/1)$.

فيمكننا إنشاء جدول مقابل للجدول (2−3) يتضمن تقاطع هذه التكاليف كما يلي:

جدول (2−3): تكاليف التصنيف الخاطئ :

المجموعات	G_1	G_2
G_1	0	$C(2/1)$
G_2	$C(1/2)$	0

ولحساب القيمة المتوقعة (المتوسط) لهذه التكاليف، نقوم بضربها بالاحتمالات الشرطية للتصنيف الخاطئ، المقابلة من الجدول (2−2) السابق، فنحصل على ما تسمى القيمة المتوقعة لتكلفة التصنيف الخاطئ (Expected Cost of misclassification)، والتي يرمز لها بـ ECM وهي تساوي حسب الجدولين (2−2) و (2−3) ما يلي:

$$\mathrm{ECM} = \mathrm{C}(2/1) * P(2/1) * P_1 + C(1/2) * P(1/2) * P_2 \qquad (12-2)$$

ويصبح هدفنا الآن أن نجعل القيمة المتوقعة ECM أصغر ما يمكن .

وبناءً على العلاقتين (2−5) و (2−6) نجد أن:

$$\mathrm{ECM} = \mathrm{C}(2/1) * P_1 \int_{R_2} f_1(x)dx + \mathrm{C}(1/2) * P_2 \int_{R_1} f_2(x)dx \qquad (13-2)$$

وبما أنه لدينا $R = R_1 \cup R_2$ فإن التكامل الشامل لـ $f_1(x)$ حسب (2−3) على المجال الكلي R يساوي الواحد ويساوي ما يلي :

$$1 = \int_R f_1(x)dx = \int_{R_1} f_1(x)dx + \int_{R_2} f_1(x)dx \qquad (14-2)$$

ومنها نجد أن تكامل $f_1(x)$ على R_2 يساوي :

$$\int_{R_2} f_1(x)dx = 1 - \int_{R_1} f_1(x)dx \qquad (15-2)$$

وبالتعويض في (2−13) نحصل على أن :

$$\mathrm{ECM} = \mathrm{C}(2/1) * P_1 \left[1 - \int_{R_1} f_1(x)dx\right] + \mathrm{C}(1/2) * P_2 \int_{R_1} f_2(x)dx$$

ومنها نحصل على أن:

$$\mathrm{ECM} = \int_{R_1} [\mathrm{C}(1/2) * P_2 f_2(x) - \mathrm{C}(2/1) * P_1 f_1(x)]dx + \mathrm{C}(2/1) * P_1 \qquad (16-2)$$

وبما أن قيم P_1 و P_2 وكذلك $C(1/2)$ و $C(2/1)$ هي أعداد موجبة، وأن جميع قيم التوزيعين $f_1(x) \geq 0$ و $f_2(x) \geq 0$ من أجل جميع قيم x في المجال R_1، فإنهما المقدران الوحيدان في ECM، اللذان يرتبطان بـ x .

وبما أن الحد الأخير $C(2/1) * P_1$ في العلاقة موجب ومعلوم، فإن ECM تأخذ أصغر قيمة لها إذا كانت نتيجة التكامل في $(2-16)$ سالبة أو معدومة .

وهكذا نجد أن ECM تأخذ أصغر قيمة ممكنة إذا كان المجال R_1 يتضمن قيماً معينة لـ x تجعل التكامل في $(2-16)$ سالباً أو معدوماً. (أي تجعل المساحة أو الحجم الذي يحدده ذلك التكامل على R_1 سالباً أو معدوماً) . وهذا يعني أن يحقق التكامل المذكور العلاقة التالية:

$$\int_{R_1} [C(1/2) * P_2 f_2(x) - C(2/1) * P_1 f_1(x)]\, dx \leq 0 \qquad (17-2)$$

وهذا لا يمكن أن يحدث إلا إذا كان المقدار المكامل (الذي تحت التكامل)سالباً أو معدوماً، أي أن ECM تأخذ أصغر قيمة ممكنة لها عندما وفقط عندما تتحقق المتراجحة التالية:

$$C(1/2) * P_2 f_2(x) - C(2/1) * P_1 f_1(x) \leq 0 \qquad (18-2)$$

وذلك من أجل قيم معينة لـ x، أي أنه باستثناء قيم x التي تجعل المقدار الأيسر موجباً. فإن المجال R_1 يجب أن يتضمن مجموعة من النقاط x التي تحقق المتراجحة $(2-18)$ والتي نكتبها على الشكل التالي:

$$C(1/2) * P_2 f_2(x) \leq C(2/1) * P_1 f_1(x) \qquad (19-2)$$

أي أن المجال R_1 يتحدد من خلال العلاقة $(2-19)$ والتي يمكن كتابتها على الشكل التالي :

$$R_1: \frac{f_1(x)}{f_2(x)} \geq \left[\frac{C(1/2)}{C(2/1)}\right]\left[\frac{P_2}{P_1}\right] \qquad (20-2)$$

كما أن المجال R_2 يجب أن يتضمن مجموعة من النقاط x التي تحقق المتراجحة المكملة التالية :

$$R_2: \frac{f_1(x)}{f_2(x)} < \left[\frac{C(1/2)}{C(2/1)}\right] * \left[\frac{P_2}{P_1}\right] \qquad (21-2)$$

ومما سبق نستنتج القاعدة التالية:

إن تحديد المنطقتين R_1 و R_2 في R، التي تجعلان ECM أصغر ما يمكن، يتم بواسطة تحديد قيم x بدلالة النسب، التي تحقق المتراجحة التالية:

$$R_1: \begin{bmatrix} \text{نسبة الكثافة} \\ \text{الاحتمالية} \end{bmatrix} \geq \left[\text{نسبة التكلفة}\right] * \begin{bmatrix} \text{معكوس نسبة} \\ \text{الاحتمالات السابقة} \end{bmatrix} \qquad \mathbf{(20-2)'}$$

$$R_2: \begin{bmatrix} \text{نسبة الكثافة} \\ \text{الاحتمالية} \end{bmatrix} < \left[\text{نسبة التكلفة}\right] * \begin{bmatrix} \text{معكوس نسبة} \\ \text{الاحتمالات السابقة} \end{bmatrix} \qquad \mathbf{(21-2)'}$$

وذلك لأن استخدام النسب أسهل من البحث عن القيم الأصلية لمركباتها .

حالات خاصة: مما سبق يمكن أن نستخلص بعض الحالات الخاصة لهذه القاعدة وهي:

أ- حالة تساوي الاحتمالات السابقة: أي عندما يكون $\frac{P_2}{P_1} = 1$ فإن القاعدة تصبح كما يلي:

$$R_1: \frac{f_1(x)}{f_2(x)} \geq \left[\frac{C(1/2)}{C(2/1)}\right] \quad , \quad R_2: \frac{f_1(x)}{f_2(x)} < \left[\frac{C(1/2)}{C(2/1)}\right] \qquad (22-2)$$

ب-حالة تساوي تكلفة التصنيف الخاطئ، أي عندما يكون $C(^1/_2) = C(^2/_1)$ فإن :

$$R_1: \frac{f_1(x)}{f_2(x)} \geq \frac{P_2}{P_1} \quad , \quad R_2: \frac{f_1(x)}{f_2(x)} < \frac{P_2}{P_1} \qquad (23-2)$$

ج- حالة تساوي الاحتمالات السابقة وتساوي التكلفة (وكذلك في حالة $\frac{C(2/1)}{C(1/2)} * \frac{P_2}{P_1} = 1$) فإن:

$$R_1: \frac{f_1(x)}{f_2(x)} \geq 1 \quad , \quad R_2: \frac{f_1(x)}{f_2(x)} < 1 \qquad (24-2)$$

ويتم استخدام هذه الحالات الخاصة في التطبيقات العملية في عدة حالات كما يلي :

إذا كانت الاحتمالات السابقة P_1 وP_2 غير معروفة فإنها تفترض على أنها متساوية، وعندها فإن عملية تصغيرECM تنحصر في مقارنة نسبة التوزيعين الاحتماليين مع نسبة تكلفة التصنيف الخاطئ (الحالة أ).

وعندما تكون تكلفة التصنيف الخاطئ غير محددة فإنها تفترض على أنها متساوية، وعندها يتم مقارنة نسبة التوزيعين الاحتمالين مع نسبة الاحتمالين السابقين $\frac{P_2}{P_1}$ (مع ملاحظة أن نسبة الاحتمالين متعاكسة مع نسبة التوزيعين) (الحالة ب) .

د- وأخيراً عندما تكون نسبتا الاحتمالين والتكلفتين متساويتين أو تساويان الواحد، أو عندما تكون أحدهما تساوي مقلوب الأخرى، فإن منطقة التصنيفR_1 تحدد ببساطة بواسطة مقارنة قيم التوزيعين مع بعضهما، وفي هذه الحالة نجد أنه إذا كانت x_0 مشاهدة جديدة وكانت:

$$R_1: \left[\left(f_1(x_0)\right) \geq f_2(x_0) \text{ أي إذا كان } \right] \qquad \frac{f_1(x_0)}{f_2(x_0)} \geq 1 \qquad (25-2)$$

فإننا ننسب x_0 إلى المجموعة G_1.

وبالمقابل إذا كانت :

$$R_2: \left[\left(f_1(x_0)\right) < f_2(x_0) \text{ أي إذا كان } \right] \qquad \frac{f_1(x_0)}{f_2(x_0)} < 1 \qquad (26-2)$$

فإننا ننسب x_0 إلى المجموعة G_2 .

وكثيراً ما يتم استخدام الحالة الأخيرة في التطبيقات العملية لأنها تسهل عمليات تصنيف العناصر والتمييز بين المجموعات، ولكنها تجعلنا نفقد الفرق بين التمييز والتصنيف (لأننا لم نعد نهتم بالحد الفاصل).

مثال (2-1): لنفترض أن التحليل الاحصائي لجميع القروض الممنوحة في المصرف حسب الدخل X أعطانا النتائج التالية:

1- إن نسبة القروض المتعثرة P_1 تساوي 30% من إجمالي القروض .

2- إن نسبة القروض غير المتعثرة P_2 تساوي 70% من إجمالي القروض .

3- إن الدخول الشهرية لأصحاب القروض تقاس بالآلاف وتتحول في المجال [75 ,25]، وإن المدرج التكراري للدخول X في مجموعة القروض المتعثرة G_1 يخضع لتوزيع بيتا $\beta_{2,4}(x)$ المعرف بالآلاف على المجال المحدد [75 ,25] والذي يعطى بالعلاقة التالية :

$$f_1(x) = \beta_{2,4}(x) = C_1(x-25)^{2-1}(75-x)^{4-1} \qquad (a)$$

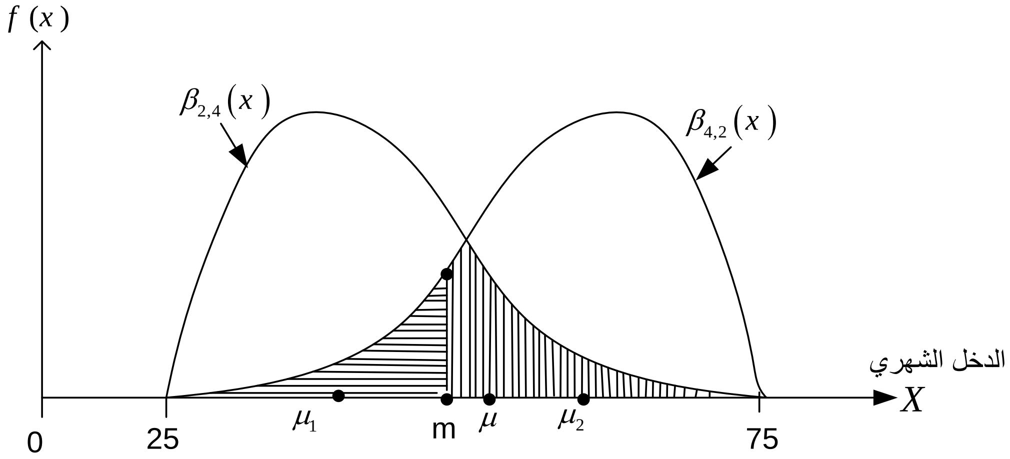
حيث: C_1 ثابت التوزيع الاحتمالي الذي يجعل المساحة تحته تساوي الواحد .

4- إن المدرج التكراري للدخول X في مجموعة القروض غير المتعثرة G_2 يخضع لتوزيع بيتا $\beta_{4,2}(x)$ المعرف بالآلاف على نفس المجال المحدد [75 ,25] والذي يعطى بالعلاقة التالية :

$$f_2(x) = \beta_{4,2}(x) = C_2(x-25)^{4-1}(75-x)^{2-1} \qquad (b)$$

حيث: C_2 ثابت التوزيع الاحتمالي الذي يجعل المساحة تحته تساوي الواحد .

5- يرسم هذان التوزيعان المنحنيين التاليين :



الشكل (2-3) **منحنيا التوزيعين** $\boldsymbol{\beta_{24}(x)}$ **و** $\boldsymbol{\beta_{42}(x)}$

ومن هذين التوزيعين يمكننا حساب توقعي وتبايني الدخل في المجموعتين μ_1 و μ_2، σ_1^2 و σ_2^2 من العلاقات التالية:

$$\mu_1 = \frac{a*q+b*p}{p+q} = \frac{25*4+75*2}{4+2} = 41.667$$

$$\sigma_1^2 = \frac{pq(b-a)^2}{(p+q)^2(p+q+1)} = 79.25 \qquad (C)$$

$$\mu_1 = \frac{a*q+b*p}{p+q} = \frac{25*2+75*4}{4+2} = 58.333$$

$$\sigma_2^2 = \frac{pq(b-a)^2}{(p+q)^2(p+q+1)} = 79.25 \qquad (e)$$

وهنا نلاحظ أن التباينين متساويان $\sigma_1^2 = \sigma_2^2 = \sigma^2 = 79.25$، وأن المتوسط العام: $\mu = 53.333$

وإن القيمة الوسطى لمتوسطيهما تساوي:

$$m = \frac{1}{2}(\mu_1 + \mu_2) = \frac{1}{2}(41.667 + 58.333) = \frac{100}{2} = 50 \qquad (f)$$

فهل تصلح هذه القيمة لتكون حداً فاصلاً بين عناصر المجموعتين G_1 و G_2 ؟

6- الآن لنفترض أن تكلفة التصنيف الخاطئ لعناصر G_1 في G_2 تساوي $C(2/1) = 20$، وإن تكلفة التصنيف الخاطئ لعناصر G_2 في G_1 تساوي $C(1/2) = 50$، وتكلفة التصنيف الصحيح معدومة .

7- مما سبق نلاحظ أن الاحتمالين السابقين P_1 و P_2 معلومان ويساويان نسبتي كل مجموعة في المجتمع، أي أن: $P_2 = 0.70$ و $P_1 = 0.30$.

وهكذا نجد أنه لتحديد مجال التمييز والتصنيف R_1 للمجموعة الأولى G_1 على المحور OX نطبق العلاقة (2-20) كما يلي :

$$R_1: \frac{f_1(x)}{f_2(x)} \geq \frac{C(1/2)}{C(2/1)} * \frac{P_2}{P_1} \geq \frac{50}{20} * \frac{0.70}{0.30} = 5.8333 \qquad (e)$$

وإن تحديد مجال التمييز R_2 للمجموعة الثانية G_2 يتم كما يلي:

$$R_2: \frac{f_1(x)}{f_2(x)} < \frac{C(1/2)}{C(2/1)} * \frac{P_2}{P_1} < \frac{50}{20} * \frac{0.70}{0.30} = 5.8333 \qquad (g)$$

وبتعويض $f_1(x)$ و $f_2(x)$ في العلاقة (e) نجد أن R_1 تحسب كمايلي:

المجال R_1 يتحدد كمايلي :

$$R_1: \frac{C_1(x-25)^1(75-x)^3}{C_2(x-25)^3(75-x)} \geq 5.8333 \qquad (h)$$

وبما أن التوزيعين $B_{2,4}(x)$ و $B_{4,2}(x)$ متناظران، فإن $(C_1 = C_2)$، وبعد إجراء الاختصارات الممكنة يكون لدينا :

$$R_1: \frac{(75-x)^2}{(x-25)^2} \geq 5.8333$$

وللحصول على الحد الفاصل نأخذ المساواة ونجذر الطرفين فنجد أن:

$$\frac{75-x}{x-25} = \pm\sqrt{5.8333} = \pm 2.415 \qquad (I)$$

$$75 - x = \pm 2.415(x - 25)$$

فإذا أخذنا الاشارة الموجبة نجد أن :

$$75 - x = 2.415x - 60.381$$

$$135.381 = 3.415x$$

$$x = \frac{135.381}{3.415} = 39.64 \qquad (k)$$

وهذه القيمة تمثل النقطة x_0 التي تمثل الحد الفاصل بين المجموعتين G_1 و G_2 .

أي أن القيمة $x = 39.64$ هي قيمة الدخل التي تفصل بين القروض المتعثرة G_1 والقروض غير المتعثرة G_2 والتي تجعل ECM أصغر ما يمكن .

وذلك ضمن شروط وخواص هذه المسألة، وهي تعطينا المجالين R_1 و R_2 المقابلين لـ G_1 و G_2 كما يلي:

$$R_1 = [25\ ,39.64] \qquad\qquad R_2 =]39.64\ \ ,75]$$

أما إذا أخذنا الإشارة السالبة للعلاقة (I) نجد أن:

$$(75 - x) = -2.415(x - 25) = -2.415x + 60.281$$

$$14.719 = -1.415x \implies x = -10.40 \qquad (\text{حل مرفوض})$$

حالات خاصة:

1-لو إننا افترضنا أن تكلفتي التصنيف الخاطئ في المجموعتين متساويتان. فإن ذلك يعني أن $\mathrm{C}(1/2) = C(2/1)$ وهذا يجعل القاعدة (2-20) تأخذ الشكل التالي:

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq \frac{P_2}{P_1} =$$

$$R_1 : \frac{C_1(x - 25)(75 - x)^3}{C_2(x - 25)^3(75 - x)} \geq \frac{0.70}{0.30} = 2.3333$$

وبعد الاختصار (ولأن $C_1 = C_2$) نحصل على أن:

$$\frac{(75 - x)^2}{(x - 25)^2} \geq 2.3333$$

وللحصول على الحد الفاصل نأخذ المساواة ونجذر الطرفين فنحصل على أن :

$$\frac{(75 - x)}{(x - 25)} = \pm 1.5275 \qquad (k)$$

وإذا أخذنا الإشارة الموجبة نحصل على أن:

$$75 - x = 1.5275(x - 25) = 1.5275x - 38.188$$
$$113.1875 = 2.5275x$$
$$x = 44.78$$

وهي تعطينا أن: $R_1 = [25\ ,44.78] \quad , \quad R_2 =]44.78\ ,75]$

أما الإشارة السالبة لـ (k) فتعطينا حلاً مرفوضاً . (تأكد من ذلك بنفسك)

2-لو افترضنا أن $\mathrm{C}(1/2) = C(2/1)$ وكانت نسبة القروض المتعثرة P_1 مساوية لنسبة القروض غير المتعثرة $(P_1 = P_2)$ فإن ذلك يجعل القاعدة (2-20) تأخذ الشكل التالي :

$$\frac{f_1(x)}{f_2(x)} \geq 1$$

$$\frac{C_1(x - 25)(75 - x)^3}{C_2(x - 25)^3(75 - x)} \geq 1$$

وبعد الاختصارات (ولأن $C_1 = C_2$) نحصل على أن:

$$\frac{(75 - x)^2}{(x - 25)^2} \geq 1$$

وللحصول على الحد الفاصل نأخذ المساواة ونجذر الطرفين فنحصل على أن:

$$\frac{75-x}{x-25}=\pm 1$$

لنأخذ الإشارة الموجبة (لأن السالبة تعطينا حلاً مرفوضاً) .

فنجد أن $75-x=x-25$

$$100=2x \;\Rightarrow\; x=50$$

وهكذا نحصل على أن الحد الفاصل بين المجموعتين G_1 و G_2 في هذه الحالة يساوي $x_0=50$ وهو يساوي القيمة الوسطى لمتوسطي المجموعتين μ_1 و μ_2 ، وهو يعطينا أن :

$$R_1=[25\ ,50]$$
$$R_2=]50\ ,75]$$

وهو يقسم المجال [75 ,25] إلى قسمين متساويين.

ملاحظة: إن نتائج هذه المسألة تختلف عن نتائج المثال العددي الأول المعروض في الجدول (2−1) لأننا أخذنا جميع عناصر المجتمع وحسبنا التوزيعين الاحتماليين للدخل في المجموعتين، ولم نكتف بعينة صغيرة منهما .

2−3 طريقة بايز Bayese (بتكاليف متساوية):

يمكن تصنيف القروض حسب دخول أصحابها حسب طريقة أخرى تسمى طريقة (بايز) أو طريقة الاحتمالات اللاحقة . وتعتمد هذه الطريقة على حساب احتمال تقاطعات الحوادث، فإذا كان لدينا حادثان B,A من جملة Ω ، فإن احتمال تقاطعهما يحسب من العلاقة التالية:

$$P(A\cap B)=P(A)*P(B/A)=P(B)*P(A/B) \qquad (27-2)$$

ومنها نحسب الاحتمالات اللاحقة $P(B/A)$ فنجد أنها تساوي:

$$P(B/A)=\frac{P(B)*P(A/B)}{P(A)} \qquad \mathbf{(28-2)}$$

ولتطبيق ذلك في التحليل التمييزي نتجاهل تكلفة التصنيف الخاطئ أو نفترضها متساوية، ونفترض أنه لدينا عنصراً عشوائياً مثل x_0 من دخول مجتمع القروض . فهو يمكن أن يكون واقعاً في أحد المجالين R_1 أو R_2 من R، ويمكن كتابته كحادث كمايلي:

$$x_0=R\cap x_0=[R_1\cup R_2]\cap x_0=[R_1\cap x_0]\cup[R_2\cap x_0] \qquad (29-2)$$

وهكذا نجد أن الاحتمال الكلي لوقوعx_0 أو في أحد المجالينR_1 وR_2، يساوي حسب (2−27) و(2−29) ما يلي:

$$P(x_0)=P(R_1)*P(x_0/R_1)+P(R_2)*P(x_0/R_2) \qquad (30-2)$$

وبما أن $P(R_1)=P_1$ و$P(R_2)=P_2$ و$P(x_0/R_1)=\not f_1(x_0)$ وأن $P(x_0/R_2)=\not f_2(x_0)$ فإن العلاقة (2−30) تأخذ الشكل التالي:

$$P(x_0)=P_1\not f_1(x_0)+P_2*\not f_2(x_0) \qquad (31-2)$$

ومن جهة أخرى نجد أن احتمال أن يقع x_0 في R_1 يساوي حسب العلاقة (2−27) مايلي:

$$\boldsymbol{P(x_0 \cap R_1) = P(x_0) * P(R_1/x_0) = P(R_1) * P(x_0/R_1) = P_1 f_1(x_0)} \quad \boldsymbol{(32-2)}$$

ومنها نجد أن الاحتمال اللاحق $P(R_1/x_0)$ يساوي:

$$P(R_1/x_0) = \frac{P_1 f_1(x_0)}{P(x_0)} = \frac{P_1 * f_1(x_0)}{P_1 f_1(x_0) + P_2 f_2(x_0)} \qquad (33-2)$$

وهو احتمال أن ينتمي العنصر المختار x_0 إلى المجال R_1، ويكون القرض المقابل له منتمياً لمجموعة القروض المتعثرة G_1 .

وبطريقة مشابهة نجد أن:

$$P(R_2/x_0) = \frac{P_2 * f_2(x_0)}{P_1 f_1(x_0) + P_2 f_2(x_0)} \qquad (34-2)$$

وهو احتمال أن ينتمي العنصر المختار x_0 إلى المجال R_2، وبذلك يكون القرض المقابل له منتمياً لمجموعة القروض غير المتعثرة G_2 .

ملاحظة: هنا يجب أن يكون:

$$P(R_1/x_0) + P(R_2/x_0) = 1 \qquad (35-2)$$

ويتم استخدام هذين الاحتمالين اللاحقين في تصنيف القروض إلى إحدى المجموعتين G_1 و G_2 حسب أكبر قيمة لهذين الاحتمالين كمايلي:

قاعدة (بايز) : إذا كان $P(R_1/x_0) > P(R_2/x_0)$، فإننا نصنف القرض المقابل في G_1

$(36-2)$ أما إذا كان $P(R_2/x_0) > P(R_1/x_0)$، فإننا نصنف القرض المقابل في G_2

مثال (2−2): لنفترض في مثالنا السابق إن دخل أحد الزبائن كان $x_0 = 40$، فكيف سنصنف القرض الذي سيحصل عليه هذا الزبون؟

نعلم من البيانات السابقة أن الاحتمالات السابقة تساوي:

$$P_1 = 0.30 \qquad\qquad P_2 = 0.70$$

ولحساب قيمتي $f_1(x_0)$ و $f_2(x_0)$ نعوض $x_0 = 40$ في قانوني التوزيع بيتا (a) و (b) فنجد أن: (علماً بأن $C_1 = C_2 = C$) :

$$f_1(x_0) = B_{2,4}(x_0) = C(40-25)(75-40)^3 = C * 643125$$

$$f_2(x_0) = B_{4,2}(x_0) = C(40-25)^3(75-40) = C * 118125$$

وبذلك نجد أن الاحتمالين اللاحقين يساويان:

$$P(R_1/x_0) = \frac{P_1 f_1(x_0)}{P_1 f_1(x_0) + P_2 f_2(x_0)} = \frac{0.3 * C * 643125}{0.3 * C * 643125 + 0.7 * C * 118125} = 0.70$$

$$P(R_2/x_0) = \frac{P_2 f_2(x_0)}{P_1 f_1(x_0) + P_2 f_2(x_0)} = \frac{0.7 * C * 118125}{0.3 * C * 643125 + 0.7 * C * 118125} = 0.30$$

وبمقارنة هذين الاحتمالين نجد أن:

$P(R_1/x_0) > P(R_2/x_0)$ لذلك فإننا نصنف القرض المقابل لذلك الدخل ضمن مجموعة القروض المتعثرة G_1. وهذا أمر يتوافق مع طريقة القيمة المتوقعة للتكاليف ECM، حيث وجدنا أنه عندما تكون التكاليف متساوية فإن الحد الفاصل التي يميز بين المجموعتين هو $\bar{x} = 44.78$ ، وبما أن $x_0 = 40 < 44.78$ فإن القرض يقع ضمن المجال R_1 المقابل للقروض المتعثرة .

ولنفترض الآن أن دخل زبون آخر كان $x_0 = 55$ فكيف سنصنف القرض الذي سيحصل عليه ذلك الزبون؟

للإجابة على ذلك السؤال نلاحظ أولاً أن قيمتي كل من الاحتمالين السابقين هي نفسها وتساوي: $P_1 = 0.30 , P_2 = 0.70$

ثم نقوم بحساب $f_1(x_0)$ و $f_2(x_0)$ فنجد أن: (لأن $C_1 = C_2 = C$):

$$f_1(x_0) = B_{24}(x_0) = C(55 - 25)(75 - 55)^3 = C * 240000$$
$$f_2(x_0) = B_{42}(x_0) = C(55 - 25)^3(75 - 55) = C * 540000$$

وبذلك نجد أن :

$$P(R_1/x_0) = \frac{0.3 * C * 240000}{0.3 * C * 240000 + 0.7 * C * 540000} = 0.16$$
$$P(R_2/x_0) = \frac{0.7 * C * 240000}{0.3 * C * 240000 + 0.7 * C * 540000} = 0.84$$

وبما أن $P(R_2/x_0) > P(R_1/x_0)$ فإننا نصنف القرض المقابل للدخل $x = 55$ ضمن مجموعة القروض غير المتعثرة G_2 .

وهذا يتفق مع طريقة القيمة المتوقعة للتكاليف ECM عندما تكون التكاليف متساوية، حيث نجد أن $x_0 > 44.78$، أي أن x_0 يقع في المجال R_2 المقابل لمجموعة القروض غير المتعثرة G_2 .

2-4 طريقة المسافة الأقرب إلى المركزين (لمتحول واحد ومجموعتين):

بعد دراسة خصائص المجموعات في المجتمع يمكننا تصنيف أي عنصر جديد مثل x_0 إلى إحدى المجموعتين بالاعتماد على قربه من مركزها . لذلك نقوم بحساب مربعي المسافة على المحور OX بين تلك النقطة ومركزي المجموعتين الأولى والثانية فنحصل على أن:

$$D_1^2 = (x_0 - \mu_1)^2 \qquad D_2^2 = (x_0 - \mu_2)^2 \qquad (37 - 2)$$

ثم نقارن بينهما ونتخذ القرار كما يلي:

إذا كانت : $D_1^2 < D_2^2$ نصنف العنصر x_0 في المجال R_1 والقرض في G_1 .

أما إذا كانت : $D_2^2 < D_1^2$ نصنف العنصر x_0 في المجال R_2 والقرض في G_2 . $(38 - 2)$

مثال (2-3): لنفترض أن دخل أحد المقترضين في المثال السابق كان $x_0 = 60$، ولتصنيفه في إحدى المجموعتين نحسب مربعي المسافة عن المركزين μ_1 و μ_2، فنجد أن :

$$D_1^2 = (60 - \mu_1)^2 = (60 - 41.666)^2 = 336.11$$
$$D_2^2 = (60 - \mu_2)^2 = (60 - 58.333)^2 = 2.772$$

ومن المقارنة نجد أن: $D_2^2 < D_1^2$، لذلك نصنف x_0 في المجال R_2 والقرض المقابل له في G_2، أي في مجموعة القروض غير المتعثرة . وهذا يتفق مع الطرائق السابقة .

2-5 طريقة تحليل التباين (لمتحول واحد ومجموعتين):

تعتمد هذه الطريقة على تحليل التباين الكلي للمتحول X في المجتمع المدروس. لذلك نعرف مجموع مربعات انحرافات قيم X عن المتوسط العام μ . ثم نقوم بتحليله إلى مجموع مربعات الانحرافات عن متوسطي المجموعتين μ_1 و μ_2 (داخل كل مجموعة)، وإلى مجموع مربعات انحرافات متوسطي المجموعتين μ_1 و μ_2 عن المتوسط العام μ كمايلي:

لذلك نرمز بـ x_{ij} للعنصر i الذي ينتمي إلى المجموعة G_j، ولنرمز لمجموع مربعات الانحرافات عن المتوسط العام μ بـ SST من (Sum Square Total) فيكون لدينا :

$$SST = \sum_{i=1}^{n_j}\sum_{j=1}^{2}\left(x_{ij}-\mu\right)^2 \qquad (39-2)$$

نطرح ونضيف متوسط المجموعة μ_j من داخل القوس ونعالجه كمايلي:

$$\begin{aligned} SST &= \sum_{i=1}^{n_j}\sum_{j=1}^{2}\left(x_{ij}-\mu_j+\mu_j-\mu\right)^2 \\ &= \sum_{i=1}^{n_j}\sum_{j=1}^{2}\left[\left(x_{ij}-\mu_j\right)^2+\left(\mu_j-\mu\right)^2-2\left(x_{ij}-\mu_j\right)\left(\mu_j-\mu\right)\right] \\ &= \sum_{i=1}^{n_j}\sum_{j=1}^{2}\left(x_{ij}-\mu_j\right)^2+\sum_{j=1}^{2}\sum_{i=1}^{n_i}\left(\mu_j-\mu\right)^2- \qquad (\mathbf{40-2}) \\ &\qquad -2\sum_{j=1}^{2}\left(\mu_j-\mu\right)\sum_{i=1}^{n_i}\left(x_{ij}-\mu_j\right) \end{aligned}$$

نلاحظ أن الحد الأخير معدوم لأن $\sum_{i=1}^{n_i}\left(x_{ij}-\mu_j\right)=0$ في كل مجموعة j، وبعد إجراء بعض الاصلاحات نحصل على أن:

$$SST = \sum_{j=1}^{2}\sum_{i=1}^{n_i}\left(x_{ij}-\mu_j\right)^2+\sum_{j=1}^{2}n_i\left(\mu_j-\mu\right)^2+0 \qquad (41-2)$$

ومنها نلاحظ أن الحد الأول هو عبارة عن مجموع مربعات انحرافات القيم x_{ij} عن متوسطات المجموعات μ_j وذلك داخل كل مجموعة. ولهذا يسمى مجموع المربعات داخل المجموعات ويرمز له بالرمز SSW من (Sum Square Within) .

أما الحد الثاني فهو عبارة عن مجموع مربعات انحرافات المتوسطين μ_1 و μ_2 عن المتوسط العام μ، لذلك يسمى مجموع المربعات بين المجموعات [Sum Square Between Groups] ويرمز له بالرمز SSB . وهكذا نجد أن SST يساوي :

$$SST = SSW + SSB \qquad (42-2)$$

نقسم الطرفين في (2–42) على SST فنحصل على العلاقة الهامة التالية :

$$\frac{SSW}{SST}+\frac{SSB}{SST}=1 \qquad (43-2)$$

والسؤال الآن كيف يمكن توظيف ذلك في تصنيف العناصر X إلى المجموعتين المذكورتين ؟

والجواب الطبيعي هو أن نجعل النسبة الداخلية $\frac{SSW}{SST}$ أصغر ما يمكن، أو أن نجعل النسبة البديلة $\frac{SSB}{SST}$ أكبر ما يمكن، وهو الأفضل .

وبكلام آخر علينا أن نصنف العناصر X داخل المجموعات بحيث تكون متجانسة داخلياً، أو بحيث نجعل مربعي الفرق بين المتوسطين μ_1 و μ_2 عن المتوسط العام μ أكبر ما يمكن. ولكن تحقيق هذا الهدف ليس بالأمر السهل ويحتاج إلى معالجات معقدة وأدوات جديدة (يمكن البحث عنها في المراجع المختصة) .

2–6 طريقة فيشر (Fisher): أو طريقة المسافة النسبية بين المركزين μ_1 و μ_2 :

انطلاقاً من أن طريقة تحليل التباين الذي يتطلب أن نجعل النسبة $\frac{SSB}{SST}$ أكبر ما يمكن، فإن (فيشر) قدم معياراً جديداً مشابهاً لتلك النسبة، وهو أن نجعل مربع المسافة بين متوسطي المجموعتين μ_1 و μ_2 مقسوماً على التباين الكلي σ_x^2 أكبر ما يمكن . أي أن نجعل المعيار التالي أكبر ما يمكن .

$$\frac{\left(\mu_1-\mu_2\right)^2}{\sigma_x^2}=\frac{D^2}{\sigma_x^2} \quad \Longrightarrow Max \qquad (44-2)$$

واعتماداً على هذه الطريقة يمكننا أن نزيح الحد الفاصل بين المجموعتين m إلى اليمين أو اليسار، بحيث نجعل المقدار (2–44) أكبر ما يمكن. لأن ذلك الانزياح سيغير من انتماء العناصر X إلى المجموعتين، وبالتالي سيغير من قيمتي μ_1 و μ_2 ومن الفرق $(\mu_1-\mu_2)$ بينهما .

مثال (2–4): لنأخذ معطيات مثال القروض السابق، ولنفترض إننا جعلنا الحد الفاصل بين الدخل في المجموعتين مساوياً لـ $x_0=50$ فعندما تتغير عناصر المجموعتين وتصبح كما يلي:

$$G_1'=\{45 \quad 29 \quad 30\} \qquad : \quad n_1=3$$

$$G_2'=\{56\ \ 56\ \ 56\ \ 56\ \ 56\ \ 56\ \ 56\ \ 64\ \ 64\ \ 72\ \ 72\ \ 72\} \ : \ n_2=12$$

وإن متوسطي هاتين المجموعتين يصبحان مساويان لـ :

$$\tilde{\mu}_1=\bar{x}_1=\frac{104}{3}=34.6667$$

$$\tilde{\mu}_2=\bar{x}_2=\frac{624}{12}=52$$

علماً بأن التباين العام لـ X يساوي $\sigma_x^2=13.153$ ويبقى ثابتاً .

وبذلك نجد أن معيار (فيشر) لمربع المسافة النسبية بينهما يساوي :

$$F=\frac{(\tilde{\mu}_1-\tilde{\mu}_2)^2}{\sigma_x^2}=\frac{(52-34.666)^2}{13.153}=\frac{300.47}{13.153}=22.844$$

وهي أكبر مما كانت عليه في التصنيف الفعلي الأول، لأن هذه النسبة حسب التصنيف الأول والذي كان حده الفاصل $x_0 = 56$ كانت تساوي:

$$F = \frac{(\tilde{\mu}_1 - \tilde{\mu}_2)^2}{\sigma_x^2} = \frac{(65.14 - 48)^2}{13.153} = \frac{293.78}{13.153} = 22.33$$

أما إذا أخذنا الحد الفاصل حسب القيمة المتوقعة للتكلفة ECM والمساوي لـ $x_0 = 39.64$ فإننا نحصل على مجموعتين جديدتين هما:

$$G_1'' = \{\, 29\,, 30\} \qquad : \quad n_2 = 2$$

$$G_2'' = \{45, 56, 56\,, 56, 56, 56, 56, 56, 64, 64, 72, 72, 72\} \quad : \; n_2 = 13$$

وإن متوسطيهما يساويان :

$$\tilde{\mu}_1 = \bar{x}_1 = \frac{59}{2} = 29.5$$

$$\tilde{\mu}_2 = \bar{x}_2 = \frac{781}{13} = 60.077$$

وهكذا نجد أن النسبة السابقة تصبح مساوية لـ :

$$F = \frac{(\tilde{\mu}_1 - \tilde{\mu}_2)^2}{\sigma_x^2} = \frac{(60.077 - 29.5)^2}{13.153} = \frac{934.95}{13.153} = 71.08$$

وهي أكبر مما كانت عليه في الحالتين المذكورتين سابقاً . وهذا يعطينا تصنيفاً للقروض ضمن مجموعتين متجانستين داخلياً ومتباينتين خارجياً، وهما المجموعتان G_1'' و G_2'' المذكورتين أعلاه، وهذا ما يجعل مربع المسافة بين مركزي هاتين المجموعتين أكبر ما يمكن .