

س) في المتغيرات التي تعبر عن تلك الخواص ، وعدد د. يقسم قيم هذه الدالة إلى جزئين بحيث إذا كانت قيمة هذه الدالة عند القياسات المشاهدة في عينة ما أصغر من العدد د. تكون العينة من المجتمع أ وإذا كانت أكبر من أو تساوى العدد د. تكون العينة من المجتمع ب ، مع بيان احتمال الخطأ في هذا التقسيم . وتسمى هذه الدالة حينئذ بدالة التمييز كما يسمى العدد د. بالنقطة الحدية boundary point .

فمثلا في حفر بئر البترول قد تكون دالة التمييز هي :

د (س_١ ، س_٢ ، ... ، س_{٢٦}) = س_١ - س_٢ + س_٣ - س_٤ + س_٥ - س_٦ + ... + س_{٢٥} - س_{٢٦} والعدد د. = ٢٦ بحيث إذا كانت قيمة هذه الدالة لقياسات أخذت على وحدات عينة ما أصغر من ٢٦ يستمر الحفر (المجتمع أ) وإذا كانت أكبر من أو تساوى ٢٦ يتوقف الحفر (المجتمع ب) . نفرض مثلا أن القيم المشاهدة للخواص الخمسة في إحدى العينات هي على الترتيب ٣٠ ، ١٢٠ ، ١٠ ، ١٦ ، ٢٠٠ . بالتعويض بهذه القيم في دالة التمييز نجد أن د (٢٠٠ ، ١٦ ، ١٠ ، ١٢٠ ، ٣٠) = ٣٨ وبما أن ٣٨ أكبر من ٢٦ تكون العينة من المجتمع ب ويتوقف الحفر .

(١٣ - ٢) إيجاد دالة التمييز :

كيف نعد الدالة د والعدد د. لتكون جاهزة للتطبيق على عينات يراد معرفة المجتمع الذى تنتمى إليه ؟ الطريق إلى ذلك ما يلى :

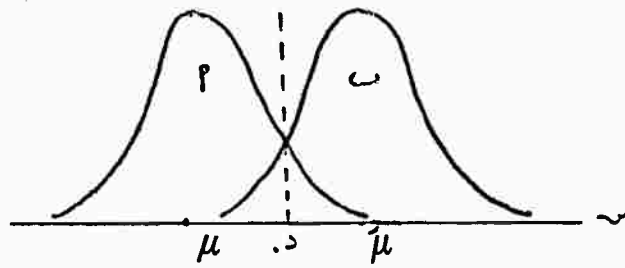
نأخذ عينة عشوائية حجمها ن_١ من المجتمع الأول وعينة عشوائية حجمها ن_٢ من المجتمع الثانى ، ونحدد ك من الخواص التى نرى أنها ذات تأثير فى التمييز بين المجتمعين (ك > ن_١ ، ك > ن_٢) ثم نقيس كل وحدة من وحدات العينتين من حيث هذه الخواص فتكون القياسات الناتجة وعددها (ن_١ + ن_٢) ك هي الأساس الذى نبني عليه دالة التمييز كما سيتبين بعد . فمثلا إذا كانت ك = ٤ تكون البيانات كما فى الجدول (١٣ - ١) الآتى :

الجدول (١٣ - ١)

عينة المجتمع الأول					عينة المجتمع الثاني				
س _١	س _٢	س _٣	س _٤		س _١	س _٢	س _٣	س _٤	
١١	٢١	٣١	٤١	١	١١	٢١	٣١	٤١	١
١٢	٢٢	٣٢	٤٢	٢	١٢	٢٢	٣٢	٤٢	٢
٠٠	٠٠	٠٠	٠٠	٠٠	٠٠	٠٠	٠٠	٠٠	٠٠
٠٠	٠٠	٠٠	٠٠	٠٠	٠٠	٠٠	٠٠	٠٠	٠٠
١,٠	٢,٠	٣,٠	٤,٠	٢,٠	١,٠	٢,٠	٣,٠	٤,٠	٢,٠
المتوسط	س _١	س _٢	س _٣	س _٤	المتوسط	س _١	س _٢	س _٣	س _٤

(أولاً) : حالة متغير واحد :

لتقديم الفكرة التي تؤسس عليها دالة التمييز ، نبدأ بالحالة التي يكون لدينا فيها متغير واحد x (ك = ١) . سنفترض أن لهذا المتغير توزيعاً معتدلاً بمتوسط μ للمجتمع الأول ، μ' للمجتمع الثاني ، أما التباين فقيمتها واحدة للمجتمعين وقدرها σ^2 . يمثل هذان التوزيعان كما في الشكل (١٣ - ١) الآتي .



الشكل (١٣ - ١)

نفرض أن s هي قيمة المتغير s التي وجدناها في عينة نعلم أنها من أحد المجتمعين ونريد تحديد المجتمع الذي تنتمي إليه العينة . إن دالة التمييز هنا تكون على الصورة $d(s) = s$. من الطبيعي أن نأخذ متوسط المجتمعين $d = \frac{1}{2}(\mu + \bar{\mu})$ كنقطة حدية تستخدم للفصل بينهما . وإذا فرضنا أن μ أصغر من $\bar{\mu}$ فإن قاعدة التمييز تكون كالآتي :

إذا كان $s > \frac{1}{2}(\mu + \bar{\mu})$ نقرر أن العينة من المجتمع أ

وإذا كان $s \leq \frac{1}{2}(\mu + \bar{\mu})$ نقرر أن العينة من المجتمع ب
ونتساءل الآن متى يكون قرارنا خاطئاً ؟ إن هذا الخطأ يقع حين تكون العينة في الواقع من المجتمع أ ويكون قرارنا أنها من المجتمع ب وهذا يحدث حيناً
(مع أن العينة من المجتمع أ) $s \leq \frac{1}{2}(\mu + \bar{\mu})$

$$\text{أي حينما } \frac{\mu - \bar{\mu}}{\sigma^2} = \frac{\mu - (\frac{1}{2}(\mu + \bar{\mu}))}{\sigma} \leq \frac{\mu - s}{\sigma}$$

$$\text{أي حينما } \frac{\delta}{\sigma^2} \leq \epsilon$$

$$(1) \quad \mu - \bar{\mu} = \delta , \quad \frac{\mu - s}{\sigma} = \epsilon \quad \text{حيث}$$

$$(2) \quad \text{واحتمال هذا الخطأ هو ل } (\frac{\delta}{\sigma^2} \leq \epsilon)$$

ويسمى هذا الاحتمال باحتمال خطأ التقسيم . ونحصل على نفس قيمة هذا الاحتمال حين تكون العينة من المجتمع ب ونقرر أنها من المجتمع أ . تحقق من ذلك .

ولما كانت ϵ تتبع التوزيع المعتدل القياسي فإننا نستطيع إيجاد هذا الاحتمال من جدول المساحات أسفل المنحنى المعتدل القياسي .

ويلاحظ أنه إذا كان المتوسطان μ ، μ متساويين أو متساويين تقريبا فإن $\frac{\delta}{\sigma \sqrt{2}}$ تكون تقريبا مساوية للصفر ويكون احتمال خطأ التقسيم مساويا بالتقريب

للنسبة ٥٠٪. ومعنى هذا أن التقسيم عشوائياً وفي هذه الحالة لا يكون هناك جدوى من إيجاد دالة صالحة للتمييز بين المحتملين. كما يلاحظ أن احتمال خطأ التقسيم يكون أصغر ما يمكن إذا كانت $\left| \frac{\mu - \mu_0}{\sigma} \right| = \left| \frac{\delta}{\sigma} \right|$ أكبر ما يمكن أى إذا كانت

القيمة الموجبة لنسبة الفرق بين متوسطي المجتمعين إلى الانحراف المعياري المشترك أكبر ما يمكن ، وتستخدم هذه الحقيقة كأساس لاشتقاق أفضل دالة للتمييز بين المجتمعين .

(ثانياً) : حالة ك من المتغيرات :

نأتى الآن للحالة التى يكون لدينا فيها ك من المتغيرات s_1 ، s_2 ، s_3 ، s_4 ، s_5 . كما سبق القول ، للتوصل إلى دالة التمييز نأخذ عينة عشوائية من كلا المجتمعين ونقيس قيم هذه المتغيرات لكل وحدة من وحدات العينتين فتكون البيانات الناتجة هى الأساس الذى نبني عليه إنشاء أفضل دالة تمييز كما يتبين بعد .

لإمكانية التحليل الاحصائى سنضع الافتراضات الثلاثة الآتية .

افتراضات التحليل :

(١) دالة التمييز خطية أى على الصورة

$$s_1 \alpha + \dots + s_r \alpha + s_{r+1} \alpha = (s) \alpha$$

(3) $\alpha \leq 1$

حيث $\alpha_1, \alpha_2, \dots, \alpha_k$ بارامترات مجهولة مطلوب تقديرها من العينتين بشرط أن نحصل من هذه التقديرات على أقل احتمال لخطأ التقسيم .

(٢) المتغيرات $x_1, x_2, \dots, x_k$ هي متغيرات عشوائية توزيعها المشترك هو توزيع معتدل متعدد (ذك من المتغيرات) سواء في المجتمع A أو في المجتمع B . وتأخذ متوسطات هذه المتغيرات القيم $\mu_1, \mu_2, \dots, \mu_k$ في المجتمع الأول ، والقيم $\bar{\mu}_1, \bar{\mu}_2, \dots, \bar{\mu}_k$ في المجتمع الثاني . أما التباين σ^2 لأي متغير x_i والذي سنرمز له بالرمز σ_i^2 فهو واحد في المجتمعين وتختلف التباينات من متغير إلى آخر ، كما أن التباين σ_i^2 ($i \neq j$) لمتغيرين x_i, x_j يكون واحدا في المجتمعين وتختلف التباينات بين كل زوج وآخر من المتغيرات .

من هذا الافتراض ينتج أن دالة التمييز المعرفة في (٣) تكون ذات توزيع معتدل لأنها خطية في متغيرات معتدلة . هذا مع تذكر أن اعتدالية التوزيع المتعدد المشترك تتضمن اعتدالية كل متغير على حدة .

(٣) العينات التي تؤخذ من كل مجتمع هي عينات عشوائية .

لايجاد أفضل تقدير لدالة التمييز المعرفة في (٣) من أزواج العينات ينبغي أن نقدر البارامترات $\alpha_1, \alpha_2, \dots, \alpha_k$ بأعداد $a_1, a_2, \dots, a_k$ تجعل الدالة $\bar{x}_i$ ذات أقل خطأ ممكن للتقسيم . وهذا الشرط كما جاء في حالة المتغير الواحد يكافئ أن تكون النسبة $\left| \frac{\delta}{\sigma} \right|$ أكبر ما يمكن حيث δ, σ هنا تعرفان كالآتي :

δ = الفرق الكلي بين المتوسطات المتناظرة في المجتمعين مرجحة بالمعاملات α_i

$$= \sum_{i=1}^k \alpha_i (\bar{\mu}_i - \mu_i)$$

$$= \sum_{i=1}^k \alpha_i \delta \quad \text{حيث } \delta = \bar{\mu}_i - \mu_i \quad (٤)$$

، $\sigma^2 =$ التباين الكلي للدالة (٣) وهي $\mu \alpha$ سـ

$$(٥) \quad \sum_{i=1}^k \sum_{j=1}^k \sigma_{ij} \alpha_i \alpha_j =$$

فمثلا إذا كانت $k = 3$ فإن

$$(\bar{\mu}_3 - \mu_3) \alpha_3 + (\bar{\mu}_2 - \mu_2) \alpha_2 + (\bar{\mu}_1 - \mu_1) \alpha_1 = \delta$$

$$\sigma_{11} \alpha_1^2 + \sigma_{12} \alpha_1 \alpha_2 + \sigma_{13} \alpha_1 \alpha_3 + \sigma_{22} \alpha_2^2 + \sigma_{23} \alpha_2 \alpha_3 + \sigma_{33} \alpha_3^2 = \sigma^2$$

$$\left| \frac{\delta}{\sigma} \right| = \Delta \text{ القيمة المطلقة } \Delta$$

أكبر ما يمكن حيث δ ، σ معرفتان في (٤)، (٥). ولتجنب الإشارات سنأخذ $\frac{\delta}{\sigma} = \Delta$ بدلا من Δ . أى أننا سنقدر البارامترات بالقيم التى تجعل الدالة

الآتية أكبر ما يمكن :

$$\frac{\delta}{\sigma} = \Delta = \frac{\sum_{i=1}^k \sum_{j=1}^k \sigma_{ij} \alpha_i \alpha_j}{\sum_{i=1}^k \sum_{j=1}^k \sigma_{ij} \alpha_i \alpha_j} = \frac{\delta}{\sigma}$$

$$(٦) \quad \frac{\sum_{i=1}^k \sum_{j=1}^k \sigma_{ij} \alpha_i \alpha_j}{\sum_{i=1}^k \sum_{j=1}^k \sigma_{ij} \alpha_i \alpha_j} = \frac{1}{2 - \rho_{12} + \rho_{13}}$$

حيث $\alpha_r =$ مجموع المربعات للمتغير x_r في المجتمعين ($r = 1, 2, \dots$)
(ك)

$\alpha_r =$ مجموع حواصل الضرب للمتغيرين x_r, x_s في المجتمعين .
وتعظيم الدالة (٦) يكافئ تعظيم الدالة

$$(٧) \quad (\alpha_r \delta_r) / (\alpha_r \alpha_s \alpha_t \dots)$$

التي لا تختلف عنها إلا في المقدار الثابت $\alpha_1 + \alpha_2 + \dots + \alpha_p$. ويقتضى هذا التعظيم إجراء التفاضل الجزئي لهذه الدالة بالنسبة إلى كل من $\alpha_1, \alpha_2, \dots, \alpha_p$ ، والمساواة بالصفر . وهذا يؤدي إلى الحصول على ك من المعادلات الخطية في ك من المجاهيل تشبه المعادلات المعتادة في الانحدار الخطي ، فهي تأخذ الصورة الآتية :

$$\delta_1 = \alpha_1 \alpha_1 + \dots + \alpha_2 \alpha_1 + \alpha_3 \alpha_1$$

$$\delta_2 = \alpha_1 \alpha_2 + \dots + \alpha_2 \alpha_2 + \alpha_3 \alpha_2$$

$$(٨) \quad \dots \dots \dots \dots \dots \dots$$

$$\delta_k = \alpha_1 \alpha_k + \dots + \alpha_2 \alpha_k + \alpha_3 \alpha_k$$

والقيم $\alpha_1, \alpha_2, \dots, \alpha_p$ ، التي تحقق هذه المعادلات جميعا تكون هي التقديرات المطلوبة للبارامترات $\alpha_1, \alpha_2, \dots, \alpha_p$ ، وبالتالي تكون دالة التمييز هي :

$$(٩) \quad \alpha_1 + \alpha_2 + \dots + \alpha_p$$

هذا مع ملاحظة أن مجاميع المربعات ومجاميع حواصل الضرب تقدر من العينتين كالآتي :

١٢ = مجموع المربعات لقيم المتغير x_1 في العينة الأولى + مجموع المربعات لقيم نفس المتغير في العينة الثانية .

وبالمثل لمجاميع المربعات الأخرى μ_{22} ، μ_{33} ، ... ، μ_k

$\sigma_1^2 =$ مجموع حواصل الضرب للمتغيرين σ_1, σ_2 في العينة الأولى + مجموع حواصل الضرب لنفس المتغيرين في العينة الثانية

وبالمثل لمجاميع حواصل الضرب الأخرى $r \neq q$

أما التباينات والتغايرات فنحصل عليها بالقسمة على درجات الحرية وهي

$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

وإيجاد هذه القيم ثم حل المعادلات (٨) يحتاج إلى الحاسب الالىكترونى توفيراً للجهد وضماناً للدقة والسرعة .

النقطة الحدية :

كما في حالة المتغير الواحد ، من الطبيعي أن نأخذ متوسط المجتمعين كنقطة حدية تفصل بينهما ، وعلى ذلك تقدر النقطة الحدية D من العينتين كالآتي :

(١٠) د. $\frac{1}{2} = \frac{1}{\sqrt{2}} (\overline{s_r} + \overline{s_r})$

وهكذا نكون قد حصلنا على دالة التمييز (٩) والقيمة الحدية (١٠) . فإذا حصلنا على قياسات S_1 ، S_2 ، ... ، S_n لعينة جديدة ، يمكن بالتعويض بهذه القياسات في الدالة (٩) ثم المقارنة بالعدد D . أن نقرر ما إذا كانت العينة تنتمي إلى المجتمع A أو إلى المجتمع B .

احتمال خطأ التقسيم PROBABILITY OF MISCLASSIFICATION

يبقى أن نقدر احتمال خطأ التقسيم لدالة التمييز التي أوجدناها . وكما في حالة المتغير الواحد تكون أكبر قيمة لهذا الاحتمال هي المعطاة بالصيغة (٢) وهي :

$$(١١) \quad \left(\frac{\delta}{\sigma} \leq \epsilon \right) \quad \text{ل}$$

حيث δ ، σ تعطيان بالصيغتين (٤) ، (٥) ، ويمكن إثبات أن $\frac{\delta}{\sigma}$ تقدر من العينتين كما يلي :

$$(١٢) \quad \frac{\delta}{\sigma} = \sqrt{\frac{1}{2} (2 - \sqrt{1 + 4\epsilon^2})} \quad \text{حيث } \bar{x}_1 - \bar{x}_2 = \delta$$

، $\bar{x}_1$ الوسط الحسابي للمتغير x_1 في العينة الأولى ،

، $\bar{x}_2$ الوسط الحسابي للمتغير x_2 في العينة الثانية .

والاحتمال (١١) يمكن إيجاده من جدول المساحات أسفل المنحنى المعتدل القياسي وهو يشير إلى درجة الثقة في دالة التمييز ، ويعتبر التقسيم جيدا بدرجة عالية إذا لم يزد هذا الاحتمال عن ٧٪ .

ملاحظات :

(١) إذا كانت دالة التمييز هي $D(x_1, x_2, \dots, x_n)$ والقيمة الحدية D_c ، وحسبنا من أزواج العينات القيمتين $D_1, D_2, \dots, D_n$ ،

د (س_١ ، س_٢ ، ... ، س_{١٠}) فينبغي أن تكون إحدى هاتين القيمتين أكبر من د. والآخرى أصغر منها .

(٢) إذا كان كل زوج من متوسطات المجتمعين متساويين (م_١ = م_٢) أو متساويين تقريبا فلا أمل في العثور على دالة تميز بين المجتمعين بكفاءة . ولهذا تجب العناية باختيار المتغيرات التي تستخدم في دوال التمييز بحيث تكون هناك فروق معقولة بين متوسطاتها .

(٣) استخدام الحاسب الآلى في هذه الدراسة أمر ضرورى لسرعة ودقة ما نريد التوصل إليه ، خاصة وأنا نضطر في كثير من الحالات إلى تجربة مجموعات مختلفة من المتغيرات لاختيار الأفضل منها . وهذا يتطلب مشقة كبيرة يغنيها عنها الحاسب .

مثال (١٣ - ١) :

في دراسة لتوزيع بكتريا النيتروجين Azotobacter في التربة كان المطلوب معرفة مدى دقة التنبؤ بوجود أو عدم وجود هذه البكتريا في التربة ، أو بمعنى آخر المطلوب إيجاد دالة خطية تميز التربة التي تحتوى على هذه البكتريا من التربة التي لا تحتوى عليها ، وذلك باستخدام ثلاث خواص كيميائية للتربة هي :

س_١ = الأس الأيدروجينى pH ، س_٢ = كمية الفوسفات الموجودة ، س_٣ = المحتوى الكلى للنيتروجين . وقد أخذت عينة عشوائية حجمها ن_١ = ١٠٠ من تربة لا تحتوى على هذه البكتريا وعينة عشوائية حجمها ن_٢ = ١٨٦ من تربة تحتوى عليها . وقيست كل وحدة من وحدات العينتين من حيث هذه الخواص الثلاث وحسبت الفروق ف_١ = س_١ - س_٢ بين أزواج المتوسطات في العينتين فوجدت كما يلي :

$$ف_١ = ٠,١٤٠٨ ، ف_٢ = ٠,٠٨٢١ ، ف_٣ = ٠,٠٨٢٦$$

وحسبت مجاميع المربعات $\sum_{i=1}^n y_i^2$ لكل من المتغيرات الثلاثة ، ومجاميع حواصل الضرب $\sum_{i=1}^n x_i y_i$ لأزواج المتغيرات فوجدت كما يلي :

$$\sum_{i=1}^n y_i^2 = 1,111, \sum_{i=1}^n x_i y_i = 1,043, \sum_{i=1}^n x_i^2 = 2,942$$

$$\sum_{i=1}^n y_i = 0,229, \sum_{i=1}^n x_i = 0,198, \sum_{i=1}^n x_i^2 = 0,051$$

وبذلك كانت المعادلات المعتادة كالآتي :

$$1,111 = 0,198 \alpha_1 + 0,229 \alpha_2 + 0,1408 \alpha_3$$

$$0,229 = 0,198 \alpha_1 + 0,051 \alpha_2 + 0,0821 \alpha_3$$

$$0,198 = 0,198 \alpha_1 + 0,051 \alpha_2 + 0,0826 \alpha_3$$

استخدم الحاسب الالكتروني في حل هذه المعادلات فأتت القيم الآتية :

$$\alpha_1 = 0,11229, \alpha_2 = 0,05310, \alpha_3 = 0,01960 \text{ وبذلك كانت دالة التمييز هي :}$$

$$D(\alpha_1, \alpha_2, \alpha_3) = 0,11229 \alpha_1 + 0,05310 \alpha_2 + 0,01960 \alpha_3$$

ويمكن حساب النقطة الحدية D من الصيغة (١٠) .

نقدر أكبر احتمال لخطأ التقسيم بهذه الدالة كما يلي :

$$\begin{aligned} \alpha_1 \alpha_2 \alpha_3 &= 0,11229 \times 0,1408 + 0,05310 \times 0,0821 + 0,0196 \times 0,0826 \\ &= 0,02179 \end{aligned}$$

$$\text{من (١٢) : قيمة } \frac{\delta}{\sigma} \text{ مقدرة من العينة} = \sqrt{0,02179 \times (2 - 186 + 100)}$$

$$= 2,4876$$

$$\text{من (١١) أكبر احتمال لخطأ التقسيم } J = \left(\frac{\delta}{\sigma} \leq \epsilon \right)$$

$$= J = (1,2438 \leq \epsilon)$$

$$= 10,75\% \text{ (من جدول المساحات أسفل المنحنى المعتدل القياسي)}$$

قد يكون من المناسب مقارنة هذا الاحتمال باحتمال خطأ التقسيم الناشئ من دالة تمييز استخدمت واحدا فقط من المتغيرات الثلاث .

نفرض أننا استخدمنا المتغير σ فقط . لدينا :

$$F = 0,1408 \text{ وهذه تقدير لقيمة } \delta$$

$$\text{الانحراف المعياري} = \sqrt{\frac{1}{2 - 186 + 100}} = \sqrt{\frac{1,111}{284}} = 0,0625$$

وهذا تقدير σ .

$$\therefore J = \left(\frac{\delta}{\sigma} \leq \epsilon \right) = \left(\frac{0,1408}{0,0625 \times 2} \leq \epsilon \right) = J = (1,126 \leq \epsilon)$$

$$= 12,92\%$$

وإذا استخدمنا المتغير $\bar{y}_p$ فقط نجد بنفس الطريقة ما يلي :

$$J = \left(\frac{\delta}{\sigma^2} \leq \epsilon \right) \quad J = (0,58 < \epsilon) = 0,28,1 \%$$

أما إذا استخدمنا المتغير $\bar{y}_p$ فقط فإن

$$J = \left(\frac{\delta}{\sigma^2} \leq \epsilon \right) \quad J = (0,41 \leq \epsilon) = 0,34,09 \%$$

ويتضح أن أفضل دالة تمييز تؤسس على متغير واحد فقط هي تلك التي تستخدم المتغير $\bar{y}_p$ ويليهما تلك التي تستخدم المتغير $\bar{y}_p$ ثم تلك التي تستخدم المتغير $\bar{y}_p$ وهذه الأخيرة تكون دالة ضعيفة للغاية لا يجوز الاعتماد عليها . على أن دالة التمييز المركبة من المتغيرات الثلاثة معا هي أفضلها جميعا .

(١٣ - ٣) اختبار تساوى أزواج المتوسطات - اختبار ت^٢ .

كما سبق الذكر في الملاحظة (٢) بالبند السابق ينبغي أن تكون المتغيرات التي تستخدم في إعداد دالة التمييز هي تلك المتغيرات الأكثر قدرة على التمييز بين المجتمعين أى التي تختلف متوسطاتها في المجتمعين اختلافا معقولا . ولذلك يهمننا في اختبار هذه المتغيرات أن نبحث دلالة الفروق بين أزواج المتوسطات في العينتين ، فإذا كانت هذه الفروق ليست ذات دلالة بمعنى أن أزواج المتوسطات في المجتمعين متساوية ، لا تكون دالة التمييز قادرة على فصل المجتمعين . أما إذا كانت تلك الفروق جوهرية فإن فرصة دالة التمييز في تقسيم المجتمعين بكفاءة تكون كبيرة .

وهناك اختبار ابتكره هوتلنج Hotelling يسمى اختبار ت^٢ يمكننا من اختبار

$$٢٢ \text{ (بين العينات) } = \frac{١٧٧٩ + ٢٨٢}{٢٨٦} = ١٣٣,١$$

$$\frac{١٨٦ \times ١٠٠}{٢٨٦} = ٦٥,٠٣٥٨٨$$

بدرجات حرية ٣

$$٠,٠٣٠٨٨ =$$

وينشأ لدينا جدول التباين الآتي .

الجدول (١٣ - ٢)

تحليل التباين لدالة التمييز - اختبارات

ط م	درجات الحرية	م م	مصدر التباين
٠,٠١٠٢٩	٣ = ك	$٠,٣٠٨٨ = \frac{١٧٧٩ + ٢٨٢}{٢٨٦} = ١٣٣,١$	بين نوعي التربة
	$١٧٧٩ + ٢٨٢$	$٠,٠٢١٧٩ =$ مح	داخل نوعي التربة
٠,٠٠٠٠٧٧	٢٨٢ = ١ - ك		

$$١٣٣,١ = ٠,٠٠٠٠٧٧ \div ٠,٠١٠٢٩ = \text{في}^{**}$$

هذه القيمة أكبر بكثير من القيمة الحرجة ف $٠,٠١ [٢٨٢, ٣]$ التي لا تزيد عن ٣,٧٨ فهي ذات دلالة عالية وتدعونا إلى رفض الفرض الصفري عن تساوي أزواج المتوسطات في المجتمعين ، وهذا ما يجب أن يكون إذا كانت دالة التمييز ذات كفاءة في فصل المجتمعين .

يلاحظ أنه إذا ظهر أن في غير ذات دلالة فإن دالة التمييز تفشل في فصل المجتمعين وينبغي حينئذ البحث عن متغيرات أخرى أو البحث عن دالة أخرى غير خطية فيما لدينا من متغيرات .

(١٣ - ٤) استخدامات دالة التمييز :

إن دالة التمييز هي وسيلة لدراسة مدى تداخل المجتمعات في بعضها أو مدى تباعدها عن بعضها . ولهذه الدالة ثلاثة أنواع من الاستخدامات تلخص فيما يلي :

(١) التقسيم والتشخيص :

يتبين هذا الهدف من المثال المقدم في بداية هذا الفصل حيث استخدمت دالة للتمييز بين المناطق التي تحتوي بترولاً والمناطق التي لا تحتوي عليه مستعينة بخمسة متغيرات جيوفيزيائية ، وكذلك من المثال (١٣ - ١) حيث أعدت دالة في ثلاثة متغيرات كيميائية تقسم التربة إلى نوعين يحتوي أحدهما على بكتريا النيتروجين ولا يحتوي الآخر عليها . كذلك إذا كان هناك نوعان من الحمى يتشابهان في الأعراض فمن المفيد أن يكتشف الطبيب القياسات الجسمية والمعملية التي تساعد على التمييز بين نوعي الحمى وأن يعرف الطريقة المثلى لضم هذه القياسات في دالة واحدة ، وكيفية تقدير درجة الثقة في تشخيص المرض .

(٢) دراسة العلاقات بين المجتمعات :

فمثلاً ، إلى أي مدى تختلف الاتجاهات والاستعدادات النفسية للمهندس الكفاء عنها في رجل الأعمال الكفاء ؟ أو هل يختلف المدخنون عن غير المدخنين اختلافاً جوهرياً في السمات النفسية والعادات السلوكية ؟

(٣) تعميم الاختبارات :

إذا أجرينا عدة قياسات على كل من عيّنتين عشوائيتين أخذتا من مجتمعين معتدلين ورغبنا في استخدام اختبار واحد للفرض الصفري عن تساوي متوسطات

هذين المجتمعين بالنسبة لجميع هذه القياسات فإن دالة التمييز تمكننا من ذلك عن طريق اختبار ت^٢ كما جاء بالبند السابق .

هذا وتجدر الإشارة إلى أن الدراسة التي قدمت في هذا الفصل عن دالة التمييز تمتد إلى الحالات التي تتناول أكثر من مجتمعين ، كما تمتد إلى الحالات التي تكون فيها دالة التمييز غير خطية .