الفصل الثالث عشر

دالة التمييز

DISCRIMINANT FUNCTION

يتناول هذا الفصل المشكلة الآتية . نفرض أن باحثا يبحث مجتمعين ا ، ب يشتركان بدرجات متفاوتة فى بعض الخواص التى يمكن قياسها عدديا . إذا كان لدى الباحث عينة عشوائية يعلم أنها من أحد هذين المجتمعين ولكنه لا يعلم ما إذا كانت من المجتمع ا أو من المجتمع ب ، وإذا كان الباحث قد قام بقياس وحدات هذه العينة من حيث ك من تلك الخواص وحصل على القيم العددية $س_1$ ، $س_2$ ، ... ، $س_ك$ فكيف يستخدم هذه القيم لتحديد المجتمع الذى تنتمى إليه العينة ؟

فمثلا إذا قامت إحدى شركات التنقيب عن البترول بحفر بئر ووجدت منطقة رملية على عمق ٤٠٠٠ قدما فكيف تقرر عن طريق قياس بعض الخواص الجيوفيزائية لعينة مأخوذة من هذه المنطقة ما إذا كانت المنطقة تختزن بترولا (المجتمع أ) فتستمر فى الحفر ، أو لا تختزن بترولا (المجتمع ب) فتتوقف عن الحفر ؟ كذلك إذا كانت إحدى الكليات تقوم بفحص الطلاب المتقدمين إليها فكيف تميز بين الطلاب الذين يصلحون للدراسة فيها (المجتمع ا) والطلاب الذين لا يصلحون لذلك (المجتمع ب) عن طريق إجراء بعض الاختبارات العلمية والنفسية على الطلاب ؟

(١٣ - ١) دالة التمييز :

إن مثل هذه المشكلات تحل عن طريق إيجاد دالة د ($س_1$ ، $س_2$ ، ... ،

$س_ن$) فى المتغيرات التى تعبر عن تلك الخواص ، وعدد $د_.$ يقسم قيم هذه الدالة إلى جزءين بحيث إذا كانت قيمة هذه الدالة عند القياسات المشاهدة فى 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 فى هذا التقسيم . وتسمى هذه الدالة حينئذ بدالة التمييز كما يسمى العدد $د_.$ بالنقطة الحدية boundary point .

فمثلا فى حفر بئر البترول قد تكون دالة التمييز هى :

$د (س_١ ، س_٢ ، ... ، س_٥) = ٢ س_١ - س_٢ + ٣ س_٣ - ٢ س_٤ + ٠,٥ س_٥$

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 فى إحدى العينات هى على الترتيب ٣٠ ، ١٢٠ ، ١٠ ، ١٦ ، ٢٠٠ . بالتعويض بهذه القيم فى دالة التمييز نجد أن $د$ (٣٠ ، ١٢٠ ، ١٠ ، ١٦ ، ٢٠٠) = 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 :

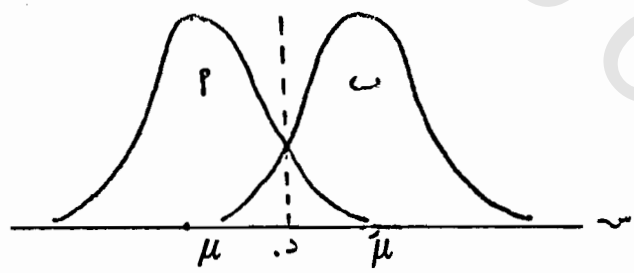
نأخذ عينة عشوائية حجمها $ن_١$ من المجتمع الأول وعينة عشوائية حجمها $ن_٢$ من المجتمع الثانى ، ونحدد ك من الخواص التى نرى أنها ذات تأثير فى التمييز بين المجتمعين ($ك < ن_١$ ، $ك < ن_٢$) ثم نقيس كل وحدة من وحدات العينتين من حيث هذه الخواص فتكون القياسات الناتجة وعددها ($ن_١ + ن_٢$) ك هى الأساس الذى نبنى 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 الثانى				
	$س_{1}$	$س_{2}$	$س_{3}$	$س_{4}$		$س'_{1}$	$س'_{2}$	$س'_{3}$	$س'_{4}$
١	$س_{11}$	$س_{21}$	$س_{31}$	$س_{41}$	١	$س'_{11}$	$س'_{21}$	$س'_{31}$	$س'_{41}$
٢	$س_{12}$	$س_{22}$	$س_{32}$	$س_{42}$	٢	$س'_{12}$	$س'_{22}$	$س'_{32}$	$س'_{42}$
..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..
$ن_{1}$	$س_{1,ن}$	$س_{2,ن}$	$س_{3,ن}$	$س_{4,ن}$	$ن_{2}$	$س'_{1,ن}$	$س'_{2,ن}$	$س'_{3,ن}$	$س'_{4,ن}$
المتوسط	$\bar{س}_{1}$	$\bar{س}_{2}$	$\bar{س}_{3}$	$\bar{س}_{4}$	المتوسط	$\bar{س}'_{1}$	$\bar{س}'_{2}$	$\bar{س}'_{3}$	$\bar{س}'_{4}$

(أولا) : حالة متغير واحد :

لتقديم الفكرة التى تؤسس عليها دالة التمييز ، نبدأ بالحالة التى يكون لدينا فيها متغير واحد س (ك = ١) . سنفترض أن لهذا المتغير توزيعا معتدلا بمتوسط μ للمجتمع الأول ، μ' للمجتمع الثانى ، أما التباين فقيمته واحدة للمجتمعين وقدرها σ^{2} . يمثل هذان التوزيعان كما فى الشكل (١٣ - ١) الآتى .



الشكل (١٣ ــ ١)

نفرض أن س هى قيمة المتغير س التى وجدناها فى عينة نعلم أنها من أحد المجتمعين ونريد تحديد المجتمع الذى تنتمى إليه العينة . إن دالة التمييز هنا تكون على الصورة د (س) = س . من الطبيعى أن نأخذ متوسط المجتمعين د. = $\frac{1}{2}(\mu + \bar{\mu})$ كنقطة حدية تستخدم للفصل بينهما . وإذا فرضنا أن μ أصغر من $\bar{\mu}$ فإن قاعدة التمييز تكون كالآتى :

إذا كان س < $\frac{1}{2}(\mu + \bar{\mu})$ نقرر أن العينة من المجتمع ا

وإذا كان س ⩾ $\frac{1}{2}(\mu + \bar{\mu})$ نقرر أن العينة من المجتمع ب

ونتساءل الآن متى يكون قرارنا خاطئا ؟ إن هذا الخطأ يقع حين تكون العينة فى الواقع من المجتمع ا ويكون قرارنا أنها من المجتمع ب وهذا يحدث حينما

س ⩾ $\frac{1}{2}(\mu + \bar{\mu})$ (مع أن العينة من المجتمع أ)

أى حينما $\frac{\mu - س}{\sigma} \geqslant \frac{\mu - \frac{1}{2}(\mu + \bar{\mu})}{\sigma} = \frac{\mu - \bar{\mu}}{\sigma 2}$

أى حينما ع ⩾ $\frac{\delta}{\sigma 2}$

حيث ع = $\frac{\mu - س}{\sigma}$ ، $\mu - \bar{\mu} = \delta$ (١)

واحتمال هذا الخطأ هو ل (ع ⩾ $\frac{\delta}{\sigma 2}$) (٢)

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

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

ويلاحظ أنه إذا كان المتوسطان μ ، $\acute{\mu}$ متساويين أو متساويين تقريبا فإن $\frac{\delta}{\sigma_{٢}}$ تكون تقريبا مساوية للصفر ويكون احتمال خطأ التقسيم مساويا بالتقريب للنسبة ٥٠٪ ومعنى هذا أن التقسيم عشوائيا وفى هذه الحالة لا يكون هناك جدوى من إيجاد دالة صالحة للتمييز بين المجتمعين . كما يلاحظ أن احتمال خطأ التقسيم يكون أصغر ما يمكن إذا كانت $\left|\frac{\mu - \acute{\mu}}{\sigma}\right| = \left|\frac{\delta}{\sigma}\right|$ أكبر ما يمكن أى إذا كانت القيمة الموجبة لنسبة الفرق بين متوسطى المجتمعين إلى الانحراف المعيارى المشترك أكبر ما يمكن ، وتستخدم هذه الحقيقة كأساس لاشتقاق أفضل دالة للتمييز بين المجتمعين .

(ثانيا) : حالة ك من المتغيرات :

:نأتى الآن للحالة التى يكون لدينا فيها ك من المتغيرات $س_{١}$ ، $س_{٢}$ ، ... ، $س_{ك}$. 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 نبنى عليه إنشاء أفضل دالة تمييز كما يتبين بعد .

لإمكانية التحليل الاحصائى سنضع الافتراضات الثلاثة الآتية .

افتراضات التحليل :

(١) دالة التمييز خطية أى على الصورة

$$د(س) = \alpha_{١} س_{١} + \alpha_{٢} س_{٢} + ... + \alpha_{ك} س_{ك}$$

$$= \sum_{ر=١}^{ك} \alpha_{ر} س_{ر} \qquad (٣)$$

حيث α_1 ، α_2 ، .... ، $\alpha_ك$ بارامترات مجهولة مطلوب تقديرها من العينتين بشرط أن نحصل من هذه التقديرات على أقل احتمال لخطأ التقسيم .

(٢) المتغيرات $س_1$ ، $س_2$ ، ... ، $س_ك$ هى متغيرات عشوائية توزيعها المشترك هو توزيع معتدل متعدد (ذو ك من المتغيرات) سواء فى المجتمع ا أو فى المجتمع ب . وتأخذ متوسطات هذه المتغيرات القيم μ_1 ، μ_2 ، ... ، $\mu_ك$ فى المجتمع الأول ، والقيم $\acute{\mu}_1$ ، $\acute{\mu}_2$ ، ... ، $\acute{\mu}_ك$ فى المجتمع الثانى . أما التباين $\sigma_ر^2$ لأى متغير $س_ر$ والذى سنرمز له بالرمز $\sigma_ر$ فهو واحد فى المجتمعين وتختلف التباينات من متغير إلى آخر ، كما أن التغاير $\sigma_{رو}$ ($ر \neq و$) لمتغيرين $س_ر$ ، $س_و$ يكون واحدا فى المجتمعين وتختلف التغايرات بين كل زوج وآخر من المتغيرات .

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

(٣) العينات التى تؤخذ من كل مجتمع هى عينات عشوائية .

لايجاد أفضل تقدير لدالة التمييز المعرفة فى (٣) من أزواج العينات ينبغى أن نقدر البارامترات α_1 ، α_2 ، ... ، $\alpha_ك$ بأعداد $ا_1$ ، $ا_2$ ، .... ، $ا_ك$ تجعل الدالة $م ا_ر س_ر$ ذات أقل خطأ ممكن للتقسيم . وهذا الشرط كما جاء فى حالة المتغير الواحد يكافىء أن تكون النسبة $\left|\frac{\delta}{\sigma}\right|$ أكبر ما يمكن حيث δ ، σ هنا تعرفان كالآتى :

δ = الفرق الكلى بين المتوسطات المتناظرة فى المجتمعين مرجحة بالمعاملات $\alpha_ر$

$$= \sum_{ر=1}^{ك} \alpha_ر (\mu_ر - \acute{\mu}_ر)$$

$$= \sum_{ر=1}^{ك} \alpha_ر \delta_ر \qquad \text{حيث } \delta_ر = \mu_ر - \acute{\mu}_ر \qquad (٤)$$

، σ^2 = التباين الكلى للدالة (٣) وهى $\sum_{ر=1}^{ك} \sum_{و=1}^{ك} \alpha_ر \alpha_و \sigma_{رو}$

$$= \sum_{ر=1}^{ك} \sum_{و=1}^{ك} \alpha_ر \alpha_و \sigma_{رو} \qquad (٥)$$

فمثلا إذا كانت ك = ٣ فإن

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

$$\sigma^2 = \alpha_1^2 \sigma_{11} + \alpha_2^2 \sigma_{22} + \alpha_3^2 \sigma_{33} + 2\alpha_1\alpha_2\sigma_{21} + 2\alpha_1\alpha_3\sigma_{31} + 2\alpha_2\alpha_3\sigma_{32}$$

المطلوب إذن إيجاد التقديرات $\alpha_ر$ بحيث تكون القيمة المطلقة $\triangle = \left| \frac{\delta}{\sigma} \right|$ أكبر ما يمكن حيث δ ، σ معرفتان فى (٤) ، (٥) . ولتجنب الإشارات سنأخذ $\triangle^2 = \frac{\delta^2}{\sigma^2}$ بدلا من $\triangle$. أى أننا سنقدر البارامترات بالقيم التى تجعل الدالة الآتية أكبر ما يمكن :

$$\triangle^2 = \frac{\delta^2}{\sigma^2} = (\sum \alpha_ر \delta_ر)^2 / \sum \sum \alpha_ر \alpha_و \sigma_{رو}$$

$$= (\sum \alpha_ر \delta_ر)^2 / \frac{1}{ن_1 + ن_2 - 2} \sum \sum \alpha_ر \alpha_و \sigma^2_{رو} \qquad (٦)$$

obeikandl.com

حيث $م_{ر ر}$ = مجموع المربعات للمتغير $س_ر$ فى المجتمعين ($ر$ = ١ ، ٢ ، ... ، ك)

، $م_{ر و}$ = مجموع حواصل الضرب للمتغيرين $س_ر$ ، $س_و$ فى المجتمعين .

وتعظيم الدالة (٦) يكافى تعظيم الدالة

$$(\text{مـ} \alpha_ر \delta_ر)^2 / \text{مـ مـ} \alpha_ر \alpha_و م_{ر و} \quad (٧)$$

التى لا تختلف عنها إلا فى المقدار الثابت $ن_1 + ن_2 - ٢$. ويقتضى هذا التعظيم إجراء التفاضل الجزئى لهذه الدالة بالنسبة إلى كل من α_1 ، α_2 ، ... ، $\alpha_ك$ والمساواة بالصفر . وهذا يؤدى إلى الحصول على ك من المعادلات الخطية فى ك من المجاهيل تشبه المعادلات المعتادة فى الانحدار الخطى ، فهى تأخذ الصورة الآتية :

$$م_{11} \alpha_1 + م_{21} \alpha_2 + ... + م_{ك1} \alpha_ك = \delta_1$$

$$م_{12} \alpha_1 + م_{22} \alpha_2 + ... + م_{ك2} \alpha_ك = \delta_2$$

$$.... \quad .... \quad .... \quad .... \quad .... \quad .... \quad (٨)$$

$$م_{1ك} \alpha_1 + م_{2ك} \alpha_2 + ... + م_{ككك} \alpha_ك = \delta_ك$$

والقيم $ا_1$ ، $ا_2$ ، ... ، $ا_ك$ التى تحقق هذه المعادلات جميعا تكون هى التقديرات المطلوبة للبارامترات α_1 ، α_2 ، ... ، $\alpha_ك$ وبالتالى تكون دالة التمييز هى :

$$ا_1 س_1 + ا_2 س_2 + ... + ا_ك س_ك \quad (٩)$$

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

$م_{11}$ = مجموع المربعات لقيم المتغير $س_1$ فى العينة الأولى + مجموع المربعات لقيم نفس المتغير فى العينة الثانية .

وبالمثل لمجاميع المربعات الأخرى $م_{22}$ ، $م_{33}$ ، ... ، $م_{kk}$

$م_{21}$ = مجموع حواصل الضرب للمتغيرين $س_1$ ، $س_2$ فى العينة الأولى + مجموع حواصل الضرب لنفس المتغيرين فى العينة الثانية

وبالمثل لمجاميع حواصل الضرب الأخرى $م_{rs}$ ($r \neq s$)

أما التباينات والتغايرات فنحصل عليها بالقسمة على درجات الحرية وهى $ن_1 + ن_2 - 2$.

وإيجاد هذه القيم ثم حل المعادلات (٨) يحتاج إلى الحاسب الالكترونى توفيرا للجهد وضمانا للدقة والسرعة .

النقطة الحدية :

كما فى حالة المتغير الواحد ، من الطبيعى أن نأخذ متوسط المجتمعين كنقطة حدية تفصل بينهما ، وعلى ذلك تقدر النقطة الحدية د. من العينتين كالآتى :

$$د. = \frac{1}{2} \sum_{r=1}^{k} ا_r (\bar{س}_r + \bar{س}'_r) \qquad (١٠)$$

وهكذا نكون قد حصلنا على دالة التمييز (٩) والقيمة الحدية (١٠) . فإذا حصلنا على قياسات $س_1$ ، $س_2$ ، ... ، $س_k$ لعينة جديدة ، يمكن بالتعويض بهذه القياسات فى الدالة (٩) ثم المقارنة بالعدد د. أن نقرر ما إذا كانت العينة تنتمى إلى المجتمع ا أو إلى المجتمع ب .

احتمال خطأ التقسيم PROBABILITY OF MISCLASSIFICATION

يبقى أن نقدر إحتمال خطأ التقسيم لدالة التمييز التى أوجدناها . وكما فى حالة المتغير الواحد تكون أكبر قيمة لهذا الاحتمال هى المعطاة بالصيغة (٢) وهى :

$$\text{ل}\left(\frac{\delta}{\sigma} \leqslant \frac{\text{ع}}{2}\right) \qquad (١١)$$

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

$$\text{تقدير } \frac{\delta}{\sigma} = \sqrt{(\text{ن}_1 + \text{ن}_2 - 2)\ \text{مح}^{\text{ك}}\ \text{أ}_\text{ر}\ \text{ف}_\text{ر}} \quad \text{حيث ف}_\text{ر} = \bar{\text{س}}_\text{ر} - \bar{\text{س}}_\text{ر} \qquad (١٢)$$

، $\bar{\text{س}}_\text{ر}$ الوسط الحسابى للمتغير $\text{س}_\text{ر}$ فى العينة الأولى

، $\bar{\text{س}}_\text{ر}$ الوسط الحسابى للمتغير $\text{س}_\text{ر}$ فى العينة الثانية .

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

ملاحظات :

(١) إذا كانت دالة التمييز هى د (س_1 ، س_2 ، ... ، $\text{س}_\text{ك}$) والقيمة الحدية $\text{د}_.$ ، وحسبنا من أزواج العينات القيمتين د ($\bar{\text{س}}_1$ ، $\bar{\text{س}}_2$ ، ... ، $\bar{\text{س}}_\text{ك}$) ،

obeikandl.com

د ($\bar{س}_1$ ، $\bar{س}_2$ ، ... ، $\bar{س}_ك$) فينبغى أن تكون إحدى هاتين القيمتين أكبر من د والأخرى أصغر منها .

(٢) إذا كان كل زوج من متوسطات المجتمعين متساويين ($\mu_1 = \mu_2$) أو متساويين تقريبا فلا أمل فى العثور على دالة تميز بين المجتمعين بكفاءة . ولهذا تجب العناية باختيار المتغيرات التى تستخدم فى دوال التمييز بحيث تكون هناك فروق معقولة بين متوسطاتها .

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

مثال (١٣ - ١) :

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

$س_1$ = الأس الأيدروجينى pH ، $س_2$ = كمية الفوسفات الموجودة ، $س_3$ = المحتوى الكلى للنيتروجين . وقد أخذت عينة عشوائية حجمها $ن_1$ = ١٠٠ من تربة لا تحتوى على هذه البكتريا وعينة عشوائية حجمها $ن_2$ = ١٨٦ من تربة تحتوى عليها . وقيست كل وحدة من وحدات العينتين من حيث هذه الخواص الثلاث وحسبت الفروق $ف_ر = \bar{س}_ر - \bar{س}_ر$ بين أزواج المتوسطات فى العينتين فوجدت كما يلى :

$ف_1$ = ٠,١٤٠٨ ، $ف_2$ = ٠,٠٨٢١ ، $ف_3$ = ٠,٠٨٢٦

وحسبت مجاميع المربعات $م_{ر ر}$ لكل من المتغيرات الثلاثة ، ومجاميع حواصل الضرب $م_{ر و}$ لأزواج المتغيرات فوجدت كما يلى :

$م_{١١}$ = ١,١١١ ، $م_{٢٢}$ = ١,٠٤٣ ، $م_{٣٣}$ = ٢,٩٤٢

، $م_{٢١}$ = ٠,٢٢٩ ، $م_{٣١}$ = ٠, ١٩٨ ، $م_{٣٢}$ = ٠,٠٥١

وبذلك كانت المعادلات المعتادة كالآتى :

١,١١١ $\alpha_{١}$ + ٠,٢٢٩ $\alpha_{٢}$ + ٠,١٩٨ $\alpha_{٣}$ = ٠,١٤٠٨

٠,٢٢٩ $\alpha_{١}$ + ١,٠٤٣ $\alpha_{٢}$ + ٠,٠٥١ $\alpha_{٣}$ = ٠,٠٨٢١

٠,١٩٨ $\alpha_{١}$ + ٠,٠٥١ $\alpha_{٢}$ + ٢,٩٤٢ $\alpha_{٣}$ = ٠,٠٨٢٦

اُستخدم الحاسب الالكترونى فى حل هذه المعادلات فأنتج القيم الآتية :

$ا_{١}$ = ٠,١١٢٢٩ ، $ا_{٢}$ = ٠,٠٥٣١٠ ، $ا_{٣}$ = ٠,٠١٩٦٠ وبذلك كانت دالة التمييز هى :

د(س$_{١}$ ، س$_{٢}$ ، س$_{٣}$) = ٠,١١٢٢٩ س$_{١}$ + ٠,٠٥٣١٠ س$_{٢}$ + ٠,٠١٩٦٠ س$_{٣}$

ويمكن حساب النقطة الحدية د$_{ح}$ من الصيغة (١٠) .

نقدر أكبر احتمال لخطأ التقسيم بهذه الدالة كما يلى :

ح$_{ا ر}$ ف$_{ر}$ = ٠,١١٢٢٩ × ٠,١٤٠٨ + ٠,٠٥٣١٠ × ٠,٠٨٢١ + ٠,٠١٩٦ ٠,٠٨٢٢٦

= ٠,٠٢١٧٩

من (١٢) : قيمة $\frac{\delta}{\sigma}$ مقدرة من العينة = $\sqrt{(١٠٠ + ١٨٦ - ٢) \times ٠,٠٢١٧٩}$

$$= ٢,٤٨٧٦$$

من (١١) أكبر احتمال لخطأ التقسيم = ل $(ع \leqslant \frac{\delta}{\sigma ٢})$

$$= \text{ل} (ع \leqslant ١,٢٤٣٨)$$

= ١٠,٧٥٪ (من جدول المساحات أسفل المنحنى المعتدل القياسى)

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

نفرض أننا استخدمنا المتغير $س_١$ فقط . لدينا :

$ف_١ = ٠,١٤٠٨$ وهذه تقدير لقيمة δ

الانحراف المعيارى = $\sqrt{\frac{١}{ن_١ + ن_٢ - ٢} م_{١١}} = \sqrt{\frac{١,١١١}{٢٨٤}} = ٠,٠٦٢٥$

وهذا تقدير σ .

$$\therefore \text{ل} (ع \leqslant \frac{\delta}{\sigma ٢}) = \text{ل}(ع \leqslant \frac{٠,١٤٠٨}{٢ \times ٠,٠٦٢٥}) = \text{ل}(ع \leqslant ١,١٢٦)$$

$$= ١٢,٩٢٪$$

obeikandl.com

وإذا استخدمنا المتغير س٢ فقط نجد بنفس الطريقة ما يلى :

$$ل\ (ع \leqslant \frac{\delta}{\sigma\ ٢}) = ل\ (ع < ٠,٥٨) = ٢٨,١٪$$

أما إذا استخدمنا المتغير س٣ فقط فإن

$$ل\ (ع \leqslant \frac{\delta}{\sigma\ ٢}) = ل\ (ع \leqslant ٠,٤١) = ٣٤,٠٩٪$$

ويتضح أن أفضل دالة تمييز تؤسس على متغير واحد فقط هى تلك التى تستخدم المتغير س١ ويليها تلك التى تستخدم المتغير س٢ ثم تلك التى تستخدم المتغير س٣ وهذه الأخيرة تكون دالة ضعيفة للغاية لا يجوز الاعتماد عليها . على أن دالة التمييز المركبة من المتغيرات الثلاثة معا هى أفضلها جميعا .

(١٣ - ٣) اختبار تساوى أزواج المتوسطات - اختبار ت٢ .

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

وهناك اختبار ابتكره هوتلنج Hotelling يسمى اختبار ت٢ يمكننا من اختبار

الفرض الصفرى المركب عن تساوى كل زوج من المتوسطات ، أى اختبار الفرض الصفرى :

$\mu_ر = \mu_{ر'}$ لجميع ر = ١ ، ٢ ، ... ، ك (١٣)

ويعتمد هذا الاختبار على تحليل التباين للمتغير مـ ا$_ر$ س$_ر$ إلى مركبتين : « بين » و« داخل » المجتمعين كالآتى :

م م (داخل العينات) = $\sum_{ر=١}^{ك} \sum_{و=١}^{ن_ر}$ ا$_ر$ ا$_و$ م$_{رو}$

= مـ ا$_ر$ ف$_ر$ بدرجات حرية $ن_١ + ن_٢ - ك = ١$ (١٤)

(تنتج الصيغة الأخيرة من ضرب المعادلات (٨) فى $\alpha_١$ ، $\alpha_٢$ ، ... ، $\alpha_ك$ ثم الجمع مع وضع ا$_ر$ بدلا من $\alpha_ر$ ، ف$_ر$ بدلا من $\delta_ر$ ٠) .

م م (بين العينات) = $\frac{ن_١ ن_٢}{ن_١ + ن_٢}$ (مـ ا$_ر$ ف$_ر$)^٢ بدرجات حرية ك (١٥)

وقد أخذت ك كدرجة الحرية بين العينات لأن التقديرات ا$_ر$ اختيرت على أساس تعظيم النسبة بين م م (بين العينات) ، م م (داخل العينات) .

فى المثال السابق نجد ما يلى :

م م (داخل العينات) = مـ ا$_ر$ ف$_ر$

= ٠,٠٢١٧٩ (سبق حسابه) بدرجات حرية ٢٨٢

obeikandl.com

م م (بين العينات) $= \frac{ن_1 ن_2}{ن_1 + ن_2} (\text{مـ}_ا \text{ ف}_ر)^2$

$= \frac{١٠٠ \times ١٨٦}{٢٨٦} (٠,٠٢١٧٩)^2$

$= ٠,٠٣٠٨٨$ بدرجات حرية ٣

وينشأ لدينا جدول التباين الآتى .

الجدول (١٣ - ٢)

تحليل التباين لدالة التميز - اختبار ت^2

مصدر التباين	م م	درجات الحرية	ط م
بين نوعى التربة	$\frac{ن_1 ن_2}{ن_1 + ن_2} (\text{مـ}_ا \text{ ف}_ر)^2 = ٠,٠٣٠٨٨$	ك = ٣	٠,٠١٠٢٩
داخل نوعى التربة	مـ$_ا$ ف$_ر$ = ٠,٠٢١٧٩	$ن_1 + ن_2$ ـ ك ـ ١ = ٢٨٢	٠,٠٠٠٠٧٧

ف$_ى$ = ٠,٠١٠٢٩ ÷ ٠,٠٠٠٠٧٧ = ١٣٣,١**

هذه القيمة أكبر بكثير من القيمة الحرجة ف$_{٠,٠١ [٣ ، ٢٨٢]}$ التى لا تزيد عن ٣,٧٨ فهى ذات دلالة عالية وتدعونا إلى رفض الفرض الصفرى عن تساوى أزواج المتوسطات فى المجتمعين ، وهذا ما يجب أن يكون إذا كانت دالة التميز ذات كفاءة فى فصل المجتمعين .

يلاحظ أنه إذا ظهر أن ف$_{ي}$ غير ذات دلالة فإن دالة التمييز تفشل فى فصل المجتمعين وينبغى حينئذ البحث عن متغيرات أخرى أو البحث عن دالة أخرىِ غير خطية فيما لدينا من متغيرات .

(١٣ – ٤) استخدامات دالة التمييز :

إن دالة التمير هى وسيلة لدراسة مدى تداخل المجتمعات فى بعضها أو مدى تباعدها عن بعضها . ولهذه الدالة ثلاثة أنواع من الاستخدامات تتلخص فيما يلى :

(١) التقسيم والتشخيص :

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

(٢) دراسة العلاقات بين المجتمعات :

فمثلا ، إلى أى مدى تختلف الاتجاهات والاستعدادات النفسية للمهندس الكفء عنها فى رجل الأعمال الكفء ؟ أو هل يختلف المدخنون عن غير المدخنين اختلافا جوهريا فى السمات النفسية والعادات السلوكية ؟

(٣) تعميم لاختبار ت :

إذا أجرينا عدة قياسات على كل من عينتين عشوائيتين أخذتا من مجتمعين معتدلين ورغبنا فى استخدام اختبار **واحد** للفرض الصفرى عن تساوى متوسطات

هذين المجتمعين بالنسبة لجميع هذه القياسات فإن دالة التمييز تمكننا من ذلك عن طريق اختبار $ت^2$ كما جاء بالبند السابق .

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

obeikandl.com