١٣ تحليل التباين ثنائى الاتجاه فى حالة عدم تساوى تكرار الفئات

Two-way analysis of variance in case of unequal subclass frequencies

١ – مقدمة

٢ – الطرق التقريبية لتحليل التباين فى حالة عدم تساوى تكرار الفئات

٣ – طريقة المتوسطات غير الموزونة

٤ – طريقة الأعداد المتوقعة للفئات

٥ – الطرق المضبوطة لتحليل التباين فى حالة عدم تساوى الأعداد فى الفئات

٦ – طريقة موائمة الثوابت للتأثيرات الثابتة للعوامل

٧ – طريقة المربعات الموزونة للمتوسطات

٨ – تنويه عن تحليل البيانات غير متناسبة التكرار

١٣-١ مقدمة

فى كثير من الحالات، يجد الباحث نفسه أمام حالات تختلف فيها أعداد المشاهدات في الفئات subclasses المختلفة المقسمة إليها البيانات ويظهر ذلك فى حالة البيانات المجمعة من السجلات، وبدرجة أقل فى حالة التجارب المخطط لها. فمثلاً فى البيانات المجمعة لدراسة ناتج الحليب من سلالة غالبا ما تكون أعداد الأبقار فى مراحل العمر وعدد مواسم الحليب تختلف من فئة إلى أخرى وأيضاً تختلف الأعداد فى المحطات المختلفة. وفى دراسات الأغنام والمعز تختلف الأعداد فى الجنسين ويختلف عدد التوائم عن الحملان المفردة ... الخ.

١٣-٢ الطرق التقريبية لتحليل التباين فى حالة عدم تساوى تكرار الفئات[1]

فى حالات كثيرة من وجود الأعداد غير المتساوية وغير المتزنة (غير المتناسبة) يرغب المجرب فى تحليل البيانات بطريقة سريعة تعطى دلالات عن تأثير العوامل المأخوذة فى الاعتبار فى النموذج الرياضى ولو أنها ليست تامة الدقة ولكنها تقريبية إلى حد ما. هذه الطرق تتميز بأنها أقل كثيراً فى التعقيد من ناحية طرق الحساب المستخدمة والتى قد تستلزم معرفة ببعض نواحى الرياضيات التى قد لا تتوافر للبعض، وفى نفس الوقت تعطى دلالة كافية عن تأثير العوامل المختلفة على مصادر الاختلافات الناشئة فى البيانات، أو أن هذه الطرق قد يتم استخدامها مبدئيا للحصول على الاتجاهات trends المعينة للبيانات انتظاراً لاستخدام الطرق الأخرى الأكثر تعقيداً فى مراحل تالية من التجربة أو عند توافر جميع البيانات، ويلاحظ أنه فى بعض الحالات قد يستوجب إجراء طرق التحليل الأكثر تعقيداً باستخدام الحاسبات الإلكترونية مما لا يتوافر فى بعض الحالات.

وهناك طريقتان تقريبيتان يمكن استخدامها فى حالة تحليل البيانات غير المتزنة وغير المتساوية وهما:

أ- طريقة المتوسطات غير الموزونة Unweighted means

ب-طريقة الأعداد الفئوية المتوقعة Expected subclass numbers

---

[1] يعتبر هذا فصل - إلى حد بعيد - تاريخياً، حيث يبين تطور هذا النوع من تحليل البيانات عندما لم يكن هناك حواسب إلكترونية ولا حتى حاسبات متقدمة. ولكنه يبين بوضوح علاقة التكرار فى الفئات باستقلال المشاهدات وتهاوى نظرية تساوى مجموع المربعات عندما تكون التكرارات غير متساوية وغير متناسبة.

وكلا الطريقتين تستخدمان نظرية التجميع الخاصة بتحليل التباين addition theorem وهى الخاصية التى تتمتع بها البيانات المتزنة (المستقلة)، وبالتالى، فإن الحسابات لمجاميع المربعات المختلفة تصبح أقل تعقيداً.

وهناك ملاحظة فى هذه النظم وهى، أنها لا تصلح عندما يكون عدد الأفراد فى أى من الفئات يساوى صفراً وسيتضح ذلك فيما بعد.

٣-١٣ طريقة المتوسطات غير الموزونة Method of unweighted means

وتعطى هذه الطريقة نتائج دقيقة إلى حد كبير لتحليل التباين إذا كانت الاختلافات بين عدد المشاهدات فى الفئات المختلفة ضيقة أى فى حدود 2:1 أو 3 ويتناقص مدى دقة precision النتائج كلما ازدادت التباينات بين أعداد المشاهدات فى الفئات المختلفة subclasses.

وطريقة المتوسطات غير الموزونة هى عبارة عن تحليل تباين لمتوسطات الفئات وبالتالى، فهى عبارة عن تحليل تباين لبيانات متساوية، حيث لكل فئة متوسط واحد، وعلى ذلك، فإن تقسيم مجموع المربعات الكلى total sum of squares إلى أجزاء خاصة بالتأثيرات الرئيسية (الأساسية) main effects وتأثير التداخل بينهما interaction يعتبر مشروعاً حيث أن نظرية التجميع لمجموع المربعات تكون صحيحة. ويلاحظ فى هذه الطريقة أن مجاميع المربعات للتأثيرات الأساسية والمحسوبة على أساس المتوسطات تختلف عن مجموع المربعات الداخلى (الخطأ) within sum of squares (error) والذى يكون محسوبا على أساس البيانات الفردية individual observations ويتم حساب مجموع المربعات للخطأ من داخل الفئات على أساس انحراف قيم المشاهدات داخل كل فئة عن متوسط هذه الفئة $\sum_i\sum_j\sum_k(Y_{ijk}-\bar{Y}_{ij.})^2$ والطريقة التى تستعمل حسابياً هى أن يحسب مجموع المربعات الكلى الغير مصحح:

$$\text{uncorrected total sum of quares} = \sum_i\sum_j\sum_k Y_{ijk}^2 \qquad (١-١٣)$$

ويحسب مجموع المربعات بين الفئات غير المصحح:

$$\text{uncorrected between subclasses sum of quares} = \sum_i\sum_j Y_{ij.}^2 \qquad (٢-١٣)$$

حيث $\sum_k Y_{ijk} = Y_{ij.}$

وبالتالى فإنه بطرح الكمية الثانية من الأولى تعطى مجموع المربعات داخل الفئات، أى

$$\sum_{i}\sum_{j}\sum_{k} Y_{ijk}^2 - \sum_{i}\sum_{j} Y_{ij.}^2$$

بمعنى أن مجموع مربعات الخطأ = مجموع المربعات الكلى – مجموع المربعات بين الفئات.

$$\text{Within SS} = \text{Total SS} - \text{Subclasses SS} \qquad (١٣-٣)$$

مما تقدم يتضح أن، متوسط المربعات داخل الفئات يتم حسابه بقسمة مجموع المربعات داخل الفئات على درجات الحرية الخاصة به والذى يساوى $(N-rc)$ حيث r تمثل عدد الصفوف (أحد التأثيرات الرئيسية)، c تمثل عدد الأعمدة (التأثير الرئيسى الآخر فى البيانات ثنائية التقسيم)، وهو بالتالى غير صحيح لاستخدامه فى اختبار متوسطات المربعات الخاصة بالتأثيرات الرئيسية أو التداخل المحسوبة على أساس متوسطات الفئات وليس المشاهدات الفردية.

ولتصحيح ذلك إما أن يصحح ليصبح متوسطات المربعات داخل الفئات محسوباً على أساس المتوسطات أو تصحيح متوسطات المربعات الخاصة بالتأثيرات الرئيسية والتداخل لتصبح محسوبة على أساس المشاهدات الفردية. وبما أن هناك متوسط مربعات واحد داخل الفئات فى مقابل عدة متوسطات مربعات للتأثيرات الأخرى، فعادة ما يتم إجراء التصحيح لمتوسط المربعات داخل الفئات. ويكون ذلك بضرب متوسطات المربعات المحسوب على أساس المشاهدات الفردية بمقلوب المتوسط التوافقى harmonic mean لعدد الأفراد فى الفئات، وبذلك يضرب متوسط المربعات داخل الفئات فى الكمية $1/n_h$ كما فى المعادلة (١٣-٤)

$$\frac{1}{n_h} = \frac{1}{rc}\left[\frac{1}{n_{11}} + \frac{1}{n_{12}} + \cdots + \frac{1}{n_{rc}}\right] \qquad (١٣-٤)$$

حيث n_{ij} تمثل عدد المشاهدات فى الفئة ij

مثال ١٣-١

تمثل بيانات جدول ١٣-١ عدد الحملان ومتوسط وزنها عند الميلاد فى دراسة عن تأثير كل من اختلاف سلالة الحمل (٤ سلالات) ونوع الولادة (مفردة أو توأمية).

جدول ١٣-١ وزن الميلاد بالكيلوجرام للحملان وأعدادها فى السلالات المختلفة

السلالة		نوع الولادة: مفردة	نوع الولادة: توأمية	المجموع
العواسى	المتوسط	4.400	3.980	**8.380**
	العدد	35	59	
النجدى	المتوسط	4.448	3.743	**8.191**
	العدد	25	35	
الدوربر	المتوسط	4.314	3.585	**7.899**
	العدد	14	26	
الحرى	المتوسط	3.300	2.713	**6.013**
	العدد	5	25	
المجموع		**16.462**	**14.021**	**30.483**

ويمكن التعبير عن قيمة المتوسط لأى من الفئات فى الجدول السابق بالنموذج الرياضى التالى:

$$\bar{Y}_{ij\cdot} = \mu + \beta_i + \tau_j + (\beta\tau)_{ij} + \bar{\xi}_{ij\cdot} \qquad (١٣-٥)$$

حيث β_i تمثل تأثير السلالة، τ_j تمثل تأثير نوع الولادة، $(\beta\tau)_{ij}$ هو التداخل بين التأثيرين على متوسط وزن الميلاد للفئة. ويشترط للحصول على الحلول ما يلى:

$$\sum_i \beta_i = \sum_j \tau_j = \sum_i (\beta\tau)_{ij} = \sum_j (\beta\tau)_{ij} = 0 \qquad (١٣-٦)$$

$$\bar{\xi}_{ij.} = \sum_k \xi_{ijk} / n_{ij} \qquad (١٣-٧)$$

كما أن هناك افتراضاً لإجراء تحليل التباين واختبار الفروض الخاصة بالتأثيرات الرئيسية وهى أن $\bar{\xi} \sim NID(0, \sigma_1^2)$. ومن خلال هذا النموذج يمكن حساب كل من مجموع المربعات الخاصة بتأثير السلالة ونوع الولادة والتداخل فيما بينهما. وهناك نموذج رياضى آخر يمكن به التعريف بمحتوى المشاهدة الواحدة وهو:

$$Y_{ijk} = \mu + \beta_i + \tau_j + (\beta\tau)_{ij} + \xi_{ijk} \qquad (١٣-٨)$$

والاشتراطات لهذا النموذج كسابقتها فى النموذج (١٣-٦) ولكن يفترض أن $\xi \sim NID(0, \sigma_e^2)$.

وهذا النموذج (١٣-٨) هو المستخدم لحساب مجموع المربعات داخل الفئات وأن σ_e^2 هى التباين للمشاهدة فى العشيرة، فى حين أن σ_I^2 هى التباين لمتوسط الفئة فى العشيرة، ولذا فإنه يتم ضرب قيمة التوقع الخاص بـ σ_e^2 بالكمية $1/n_h$ للحصول على التوقع الخاص بـ σ_I^2 وبالتالى يمكن إجراء اختبارات المعنوية. وتكون الحسابات المطلوبة لإجراء تحليل التباين بطريقة المتوسطات غير الموزونة كما يلى:

١- من حسابات سابقة مبنية على أساس المشاهدات الفردية وجد أن قيمة متوسط المربعات داخل الفئات (within MS) يساوى 0.285

٢- معامل التصحيح CF (من جدول ١٣-١): $(30.483)^2/8 = 116.152$

٣- مجموع المربعات بين الفئات Subclasses SS:

$$(4.4)^2 + (3.98)^2 + \cdots + (2.713)^2 - CF = 2.556$$

٤- مجموع المربعات للسلالات Breeds SS:

$$[(8.38)^2 + (8.191)^2 + (7.899)^2 + (6.013)^2]/2 - CF = 1.782$$

٥- مجموع المربعات لنوع الولادة Type of birth SS:

$$[(16.462)^2 + (14.021)^2]/4 - CF = 0.7448$$

٦- مجموع المربعات للتداخل Interaction SS:
= مجموع المربعات للفئات- (مجموع المربعات للسلالات + مجموع المربعات لنوع الولادة) = 0.0299

٧- مقلوب المتوسط التوافقى: $\frac{1}{n_h} = \frac{1}{8}\left[\frac{1}{35} + \frac{1}{59} + \ldots + \frac{1}{25}\right] = 0.05821$

وهنا يتضح، ضرورة أن تكون كل الفئات بها فرد واحد على الأقل لكى يمكن حساب المتوسط التوافقى، أى تجنب أن يكون مقام أحد هذه الكسور صفر.

٨- متوسط المربعات للخطأ = (متوسط المربعات داخل الفئات) $(1/n_h)$:

$$(0.285)(0.05821) = 0.01659$$

وبالتالى، يمكن استكمال البيانات فى جدول ١٣-٢ لتحليل التباين للمتوسطات لغير موزونة إذا افترض أن تأثير كل من السلالات ونوع الولادة ذات تأثير ثابت fixed effect.

جدول ١٣-٢ تحليل التباين لبيانات جدول ١٣-١ باستخدام طريقة المتوسطات غير الموزونة

SOV	df	SS	MS	F
السلالات (B)	3	1.7820	0.59	35.54**
نوع الولادة (T)	1	0.7448	0.74	44.58**
التداخل (B x T)	3	0.0299	0.01	0.60^{NS}
Error الخطأ	214		0.0166	
المجموع	221			

ولاختبار الفروض الخاصة بتأثير كل من السلالة ونوع الولادة فى النموذج الرياضى الذى تم افتراضه، وفى حالة التأثيرات الثابتة للعوامل fixed effects فإن متوسط المربعات الخاصة بكل من التأثيرات الرئيسية والتداخل يختبر مقابل متوسط المربعات للخطأ.

وكما سبق إيضاحه فى فصل ١١-٥ بأنه إذا كانت التأثيرات الرئيسية ذات تأثير عشوائى، فإنه لاختبار متوسط المربعات لها يكون باستخدام متوسط المربعات الخاص بالتداخل بين العاملين. أما إذا كان أحد التأثيرات عشوائيا والآخر ثابتا، فيتم اختبار متوسط المربعات الخاص بالعامل العشوائى التأثير بمقارنته بمتوسط مربعات الخطأ، أما العامل الثابت فيختبر مقابل متوسط المربعات للتفاعل. وفى جميع الأحوال يختبر متوسط المربعات للتداخل بين العوامل مع متوسط مربعات الخطأ، ويجدر الملاحظة، أنه فى معظم التجارب يود الباحث معرفة تباين القيمة الفردية فى مثل هذه التجارب، هذه القيمة لا يمكن التعرف عليها من الجدول إلا إذا ذكرت قيمة n_h.

١٣-٤ طريقة الأعداد المتوقعة للفئات
Method of expected subclass numbers

يلاحظ فى كثير من التجارب التى تجرى أن أعداد المشاهدات فى الفئات المختلفة تكون غير متناسبة disproportionate، فى حين يعتقد المجرب أن عنده من الأسباب

ما يقنعه بأن الأعداد الخاصة بالفئات المختلفة في العشيرة التى أخذت منها العينة (التجربة) تتمتع بتناسب فى الأعداد فيما بينها. فمثلاً قد تنطبق نسبة الذكور إلى الإناث على ما هو معروف من تساوى نسبة الجنسين فى العشيرة، أو أن عدد الحملان التوأمية والمفردة التى يحصل عليها المجرب فى بياناته، لا تنطبق عليها النسبة المعروفة فى إحدى سلالات الأغنام، وبناء على ذلك، فيكون السبب فى اختلاف النسب المشاهدة للفئات المختلفة فى العينة عنها فى العشيرة يرجع إلى المعاينة sampling variation أى اختلافات العينات عن بعضها وعن نسب العشيرة، ففى هذه الحالة يمكن استخدام طريقة التحليل المبنية على أساس الأعداد المتوقعة للفئات كما هى فى العشيرة. ويكون ذلك بأن تعدل البيانات (الأعداد أو النسب للفئات المختلفة) لتصبح متناسبة وبالتالى، فإن ما حدث من عدم التطابق لنظرية التجميع additivity لمجاميع المربعات يعود ثانية، وتصبح مجاميع المربعات تجميعية فى تحليل التباين.

ومبدئياً يلزم أن يختبر المجرب ما إذا كان الافتراض بأن الأعداد فى الفئات المختلفة مأخوذة من العشيرة التى بها أعداد الفئات متناسبة من عدمه ويكون ذلك بإجراء اختبار مربع كاى χ^2 test (الباب السابع). فإذا افترض أن عدد المشاهدات فى الفئة التابعة للصف i، والعمود j هى n_{ij} فإن القيمة المتوقعة $E(n_{ij})$ بناء على النسب الموجودة فى العشيرة يمكن حسابها كالتالى:

$$E(n_{ij}) = (n_{i.})(n_{.j})/n_{..} \qquad (٩-١٣)$$

حيث $n_{i.}$ هو مجموع المشاهدات فى الصف i ، $n_{.j}$ هو مجموع المشاهدات فى العمود j ، $n_{..}$ هو العدد الكلى للمشاهدات فى التجربة. والقيمة $E(n_{ij})$ هى ما يتوقعه المجرب للفئة ij إذا كانت ينطبق عليها نفس النسب الموجودة فى العشيرة الأصلية. وبالتالى، فإن قيمة χ^2 التى تختبر ما إذا كانت العينات فعلاً متناسبة تحسب على أساس:

$$\chi^2 = \sum_i \sum_j \frac{[n_{ij} - E(n_{ij})]^2}{E(n_{ij})} \qquad (١٠-١٣)$$

وهى الصورة العامة لاختبار مربع كاى.

وتتم مقارنة القيمة المتحصل عليها لمربع كاى χ^2 بالقيمة الجدولية عند مستوى المعنوية المفترض، وبدرجات حرية تساوى $(r-1)(c-1)$ حيث c, r هما عدد الصفوف والأعمدة على التوالى. فإذا كانت قيمة χ^2 المحسوبة معنوية فإن ذلك يدل على أن العينة مأخوذة من عشيرة أعداد الفئات فيها ليست متناسبة. أى أن السبب فى

كون الأعداد فى العينة غير متناسبة لا يرجع إلى التباين فى العينات ولكن لطبيعة العشيرة نفسها.

وبناء على ذلك فإن تحليل التباين لا يمكن إجراؤه بهذه الطريقة. وعموماً فإن اختبار χ^2 يعتبر اختباراً مبدئياً. أى أنه سوف يكون له بعض التأثير على مستوى المعنوية التى يتم بها اختبار كل من التأثيرات الرئيسية والتداخل لاحقاً، ولو أنه يمكن، إلى حد ما، عدم أخذ هذا التأثير فى الاعتبار. ولإجراء تحليل التباين باستخدام طريقة الأعداد المتوقعة للفئات بفرض أن قيمة اختبار χ^2 المبدئية كانت غير معنوية فإنه يتم حساب لكل فئة من الفئات subclass مجموعاً total جديداً على أساس القيمة المحسوبة للعدد المتوقع لهذه الفئة والتى عرفت على أساس $E(n_{ij})$ بضربها فى قيمة المتوسط الحقيقى المحسوب لهذه الفئة، وبالتالى يكون هناك تكرار أو عدد جديد لكل فئة هو قيمة $E(n_{ij})$ ومجموعا جديدا يساوى $\overline{Y}_{ij}[E(n_{ij})]$، أما بالنسبة لمجموع المربعات الداخلى within sum of squares فإنه يتم حسابه كما سبق فى (١٣-٣) أى من المشاهدات الأصلية داخل الفئات ويتم بعد ذلك إجراء حساب مجاميع المربعات للتأثيرات وللتداخل بالطريقة المتبعة عادة باستخدام المجاميع للفئات المعدلة للنسب.

مثال ١٣-٢

يوضح جدول ١٣-٣ بيانات أوزان الحملان من الذكور والإناث لأربعة سلالات مختلفة عند عمر حوالى 170 يوما.

وكما ذكر فإنه يتم إجراء اختبار مبدئى لمعرفة ما إذا كانت الأعداد المتحصل عليها فى التجربة تتبع حقيقة العشيرة التى بها أعداد متناسبة للفئات ويتم ذلك بحساب الأعداد المتوقعة $E(n_{ij})$ كما فى معادلة (١٣-٩).

فمثلاً: $21.23 = (69)(44)/143$، $(74)(32)/143 = 16.56$ ... وهكذا.

ومن هذه الأعداد يتم حساب قيمة χ^2 كما فى المعادلة (١٣-١٠)

$$\chi^2 = \frac{(19-16.56)^2}{16.56} + \frac{(13-15.44)^2}{15.44} + \cdots + \frac{(22-19.78)^2}{19.78} = 1.69$$

بدرجات حرية $3 = (2-1)(4-1)$

وعند مقارنة قيمة χ^2 المحسوبة بقيمة χ^2 من جدول ٦ ملحق أ عند مستوى معنوية 5% أى $[\chi^2_{(0.05,3)} = 7.81]$ يتضح أنها غير معنوية.

جدول ١٣-٣ أوزان الحملان من أربعة سلالات مقسمة حسب الجنس لحسابات طريقة الأعداد المتوقعة للفئات

السلالة		الجنس: إناث – المشاهد	إناث – النسبى	ذكور – المشاهد	ذكور – النسبى	المجموع
الدوربر	n_{ij}	19	16.56	13	15.44	**32**
	$\overline{Y}_{ij}$	30.37		27.88		
	$Y_{ij\cdot}$		502.93		430.47	**933.40**
العواسى	n_{ij}	24	22.77	20	21.23	**44**
	$\overline{Y}_{ij}$	27.00		30.15		
	$Y_{ij\cdot}$		614.79		640.08	**1254.87**
الحرى	n_{ij}	12	13.45	14	12.55	**26**
	$\overline{Y}_{ij}$	16.71		18.14		
	$Y_{ij\cdot}$		224.75		227.66	**452.41**
النجدى	n_{ij}	19	21.22	22	19.78	**41**
	$\overline{Y}_{ij}$	29.63		26.32		
	$Y_{ij\cdot}$		628.75		520.61	**1149.36**
المجموع	$\sum_i n_{ij}$	**74**		**69**		**143**
	$\sum_i Y_{ij}$		**1971.22**		**1818.82**	**3790.04**

وفى جدول ١٣-٣ تحسب $\overline{Y}_{ij} = Y_{ij\cdot}$ مضروبة فى العدد النسبى $E(n_{ij})$ أى مثلاً 502.93 فى الخلية الأولى = (30.37)(16.56) وأن 520.61 فى الخلية الأخيرة = (26.32)(19.78) ... وهكذا.

وبناء على ذلك فإنه يتم استكمال تحليل التباين باستخدام هذه الطريقة، وكما سبق فإنه يتم حساب مجموع مربعات الخطأ من البيانات الأصلية من داخل الفئات subclasses والغير مذكورة فى جدول ١٣-٣.

مجموع مربعات الخطأ: within SS = 2721.1852

ويتم بعد ذلك استخدام البيانات الموجودة فى جدول ١٣-٣ لاستكمال تحليل التباين باستخدام الأعداد المتوقعة والمجاميع الجديدة المحسوبة وهى كالتالى:

معامل التصحيح:

$$CF = \frac{(3790.04)^2}{143} = 100450.4$$

مجموع المربعات بين الفئات Subclasses SS:

$$\text{Subclasses SS} = \frac{(502.923)^2}{16.56} + \cdots + \frac{(520.6)^2}{19.78} - CF = 2940.6$$

مجموع المربعات بين السلالات (B) SS :

$$SS\ (B) = \frac{(933.39)^2}{32} + \cdots + \frac{(1149.36)^2}{41} - CF = 2656.6$$

مجموع المربعات بين الجنسين (S) SS :

$$SS\ (S) = \frac{(1971.22)^2}{72} + \frac{(1818.82)^2}{69} - CF = 2.7$$

مجموع السربعات للتداخل (S x B) SS :

$$SS\ (S \times B) = 2940.7 - (2656.71 + 2.77) = 281.3$$

ويتم عرض النتائج المتحصل عليها فى جدول تحليل التباين (جدول ١٣-٤) التالى:

جدول ١٣-٤ تحليل التباين للبيانات فى مثال ١٣-٢ باستخدام طريقة الأعداد المتوقعة للفئات

SOV	df	SS	MS
بين السلالات	3	2656.6	885.5**
بين الجنسين	1	2.7	2.7
التداخل	3	281.3	93.8**
الخطأ(أو داخل الفئات)	135	2721.2	20.16
المجموع	142		

وباختبار الفرض الخاص بمعنوية التداخل بين العاملين بقسمة متوسط المربعات الخاص بالتداخل على الخطأ وجد أنه معنوى جداً وهو اختبار مباشر. أما بالنسبة لاختبار التأثيرات الرئيسية فإنه نتيجة لعدم تساوى الأعداد تختلف قيم المعاملات لكل من مكونات التباين ولقد أوضح Searle (1986) مكونات كل من متوسطات المربعات فى هذه الحالة حسب ما إذا كانت العوامل الرئيسية فى النموذج ثابتة fixed أو عشوائية random أو مختلطة mixed وبالتالى يمكن تكوين المقام الملائم لإجراء اختبار F حسب النموذج الرياضى والعامل الرئيسى الذى يجرى اختباره.

وفى التجربة السابقة إذا كان النموذج الرياضى المستخدم يفترض أن تأثير العوامل الرئيسية ثابت فعليه يمكن إجراء اختبار F باستخدام متوسط المربعات للخطأ ولو أنه يعتبر اختباراً غير دقيق. ونتيجة لكون التداخل بين العوامل تأثيرها معنوياً فإنه لا يصبح هناك معنى كبير للآثار الرئيسية، ولكن قد ينظر إلى حجم تأثير كل من العاملين ومشاركته فى مجموع المربعات. ويتضح هذا فى اختلاف حجم التأثير للعامل الأول مقارنة بالعامل الثانى.

صندوق ١٣-١

إذا لم يؤخذ فى الاعتبار عدم تناسب أعداد المشاهدات (الوحدات التجريبية) فى الفئات فإنه يتسبب عنها مشكلتان هما: عدم استقلالية المشاهدات وخرق لقاعدة مجموع المربعات. ولتفادى هذا ذكرت طريقتان تقريبيتان هما طريقة المتوسطات غير الموزونة وطريقة الأعداد المتوقعة للفئات. وكلتا الطريقتين يحققان هدف تجاوز المشكلتين السابق ذكرهما. ولكنهما طريقتين تقريبيتين والثانية منهما تتطلب فروضاً معينة مثل أن عدم التناسب فى الأعداد مرجعه عشوائى وليس بسبب معاينة غير موفقة أو بسبب التجربة نفسها (مثل نفوق حيوانات نتيجة للمعاملة مثلاً).

١٣-٥ الطرق المضبوطة لتحليل التباين فى حالة عدم تساوى الأعداد فى الفئات

Exact methods for analysis of variance for data with unequal subclass frequencies

إن طرق تحليل التباين المضبوطة للبيانات التى تتصف بعدم تساوى الأعداد فى فئاتها المختلفة تحتاج إلى حسابات معقدة بعض الشئ، ويحتاج بعضها إلى المعرفة بنواحى معينة من الرياضيات مثل جبر المصفوفات، خاصة عندما يزداد عدد التقسيمات classifications أو العوامل factors وبالتالى تتفاوت الأعداد فى الفئات المختلفة تفاوتاً كبيراً، كما وأن هناك طرق تحليل مضبوطة تتبع فى حالة احتواء النموذج الرياضى على تداخلات بين العوامل أو فى حالة عدم وجودها.

وسوف يتم استعراض بعض الطرق المستخدمة فى التحليل فى حالة وجود عاملين أو أساسين للتقسيم أى حالة two-way classifications (والتى يمكن توسيع مداها لتشمل حالات وجود أكثر من عاملين) ففى هذه الحالة ينشأ اتجاهين للتقسيم أى عاملان A، B ولكل منهما عدد من المستويات، فإذا افترض أن عدد المستويات لعامل الأول A تساوى r أى يصبح هناك $A_1, A_2, \cdots, A_r$ وأن عدد المستويات لعامل الثانى B تساوى c أى أن هناك $B_1, B_2, \cdots, B_c$ فيكون العدد الكلى للفئات subclasses هو rc وهو حاصل ضرب مستويات A_i فى مستويات B_j. أما بالنسبة لعدد المشاهدات فى كل فئة فيختلف، وبذلك تكون قيم n_{ij} مختلفة. وبذا ينشأ الشكل التالى للبيانات كما فى جدول ١٣-٥.

حيث تكون التعويضات الموجودة فى الجدول كالتالى:

عدد المشاهدات فى الفئة ij: n_{ij}

مجموع قيم المشاهدات فى الفئة ij: $Y_{ij.} = \sum_k Y_{ijk}$

قيمة المشاهدة k فى الفئة ij: Y_{ijk}

عدد المشاهدات فى المستوى A_i: $n_{i.} = \sum_j n_{ij}$

مجموع قيم المشاهدات فى المستوى A_i: $Y_{i..} = \sum_j \sum_k Y_{ijk}$

عدد المشاهدات فى المستوى B_j: $n_{.j} = \sum_i n_{ij}$

جدول ١٣-٥ التقسيم الثنائي في حالة اختلاف الأعداد الفئوية

		B					
		B_1	B_2	...	B_j	...	B_c	**Total**
A	A_1	$Y_{11\cdot}$	$Y_{12\cdot}$	...	$Y_{1j\cdot}$	...	$Y_{1c\cdot}$	$Y_{1\cdot\cdot}$
		n_{11}	n_{12}	...	n_{1j}	...	n_{1c}	$n_{1\cdot}$
	A_2	$Y_{21\cdot}$	$Y_{22\cdot}$	...	$Y_{2j\cdot}$	...	$Y_{2c\cdot}$	$Y_{2\cdot\cdot}$
		n_{21}	n_{22}	...	n_{2j}	...	n_{2c}	$n_{2\cdot}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	A_i	$Y_{i1\cdot}$	$Y_{i2\cdot}$	...	$Y_{ij\cdot}$	...	$Y_{ic\cdot}$	$Y_{i\cdot\cdot}$
		n_{i1}	n_{i2}	...	n_{ij}	...	n_{ic}	$n_{i\cdot}$
	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	A_r	$Y_{r1\cdot}$	$Y_{r2\cdot}$	...	$Y_{rj\cdot}$	...	$Y_{rc\cdot}$	$Y_{r\cdot\cdot}$
		n_{r1}	n_{r2}	...	n_{rj}	...	n_{rc}	$n_{r\cdot}$
Total		$Y_{\cdot 1\cdot}$	$Y_{\cdot 2\cdot}$	...	$Y_{\cdot j\cdot}$	...	$Y_{\cdot c\cdot}$	$Y_{\cdots}$
		$n_{\cdot 1}$	$n_{\cdot 2}$	...	$n_{\cdot j}$	...	$n_{\cdot c}$	$n_{\cdot\cdot}$

مجموع قيم المشاهدات في المستوى B_j: $$Y_{\cdot j\cdot} = \sum_i \sum_k Y_{ijk}$$

عدد المشاهدات الكلية: $$n_{\cdot\cdot} = \sum_i \sum_j n_{ij} = \sum_i n_{i\cdot} = \sum_j n_{\cdot j}$$

مجموع قيم المشاهدات الكلية: $$Y_{\cdots} = \sum_i \sum_j \sum_k Y_{ijk}$$

وبالطبع يمكن الحصول على متوسط كل فئة أو كل مستوى بقسمة المجموع الخاص بالفئة أو المستوى على عدد المشاهدات لهذا المجموع.

وسوف يتم التطرق إلى الطرق المستخدمة في تحليل التباين لهذا النوع من البيانات.

١٣-٦ طريقة موائمة الثوابت للتأثيرات الثابتة للعوامل

Method of fitting constants for fixed effects

تستخدم هذه الطريقة فى الحالات التى لا يفترض النموذج الرياضى فيها وجود التداخل بين العوامل المؤثرة على تباين قيم المشاهدات بل فقط التأثيرات الرئيسية main effects، ولا يلغى هذا إمكانية حساب التداخل من نتائج تحليل العينة وبالتالى اختبارها. ويعرف هذا النوع من التحليل أيضاً بطريقة المربعات الصغرى least squares analysis. وهذه الطريقة فى التحليل للبيانات غير المتساوية وغير المتناسبة أكثر الطرق المضبوطة استخداماً ويمكن لتسهيل العمليات الحسابية البدء بالتحليل المبدئى للبيانات ويكون كالتالى:

١- يحسب مجموع المربعات للفئات Subclasses SS وهو

$$\sum_i \sum_j Y_{ij.}^2 / n_{ij} - CF \qquad (١٣-١١)$$

حيث معامل التصحيح هو $CF = Y_{...}^2 / n..$. ومجموع المربعات هذا له درجات حرية تساوى $(rc-1)$.

٢- يحسب مجموع المربعات الخاص بتأثير العامل الأول A وبإهمال أى تأثير للعامل الثانى B أى $(A, \text{ignoring } B)$ وله $(r-1)$ درجات حرية

$$SS\ A, \text{ignoring } B = \sum_i \frac{Y_{i..}^2}{n_{i.}} - CF \qquad (١٣-١٢)$$

٣- بنفس الطريقة يحسب مجموع المربعات الخاص بتأثير العامل الثانى مع إهمال تأثير العامل الأول $(B, \text{ignoring } A)$ وله $(c-1)$ درجات حرية

$$SS\ B, \text{ignoring } A = \sum_j \frac{Y_{.j.}^2}{n_{.j}} - CF \qquad (١٣-١٣)$$

٤- أخيراً يتم حساب مجموع المربعات داخل الفئات within subclasses SS وهذه الكمية لها درجات حرية تساوى $(n..-rc)$

$$\text{within subclasses SS} = \sum_i \sum_j \sum_k Y_{ijk}^2 - \sum_i \sum_j \frac{Y_{ij.}^2}{n_{ij}} \qquad (١٤-١٣)$$

أو يطرح مجموع المربعات للفئات من مجموع المربعات الكلى بعد تصحيح كليهما.

وبالطبع فإن مجموع المربعات الكلى $\text{Total SS} = \sum_i \sum_j \sum_k Y_{ijk}^2 - CF$

وله درجات حرية تساوى $(n_{..} - 1)$

وكما سبق فإن النموذج الرياضى المستخدم فى هذا التحليل يكون كالتالى:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \xi_{ijk} \qquad (١٥-١٣)$$

حيث $i = 1, 2, \cdots, r$ ، $j = 1, 2, \cdots, c$ ، $k = 1, 2, \cdots, n$

ويفترض حتى يمكن إجراء اختبارات المعنوية أن $\xi \sim NID(0, \sigma^2)$. ولإيجاد حل لتأثير العوامل، أو بمعنى آخر حتى يمكن حل المعادلات الاعتيادية normal equations والتى تنتج من جعل مجموع المربعات للانحرافات أقل ما يمكن يجب وضع الاشتراط التالى:

$$\sum_i \alpha_i = \sum_j \beta_j = 0 \qquad (١٦-١٣)$$

أو أى اشتراط آخر يمكن من حل المعادلات الطبيعية (مثلاً $\alpha_1 = \beta_1 = 0$).

وبناء على نظرية أقل مجموع مربعات theory of least squares فإنه للحصول على التقديرات للثوابت التى تعطى أقل مجموع مربعات للخطأ أن يتم إجراء التفاضل للنموذج الرياضى (١٣-٥) أى أن المطلوب هو الحصول على تقديرات للثوابت تجعل الكمية

$$\sum_{ijk} e_{ijk}^2 = \sum [Y_{ijk} - \mu - \alpha_i - \beta_j]^2 \qquad (١٧-١٣)$$

أقل ما يمكن.

ويمكن الحصول على المعادلات الاعتيادية normal equations بإجراء التفاضل الجزئى بالنسبة للمعالم ومساواة المعادلة بالصفر والتى تعطى الحلول أو التقديرات

للثوابت بعد تطبيق الاشتراطات السابق ذكرها والتى يمكن تلخيصها فى جدول ١٣-٦

وبالنظر إلى المعادلات الاعتيادية يتضح أن المعامل coefficient لكل من الثوابت فى معادلة رقم 1 يساوى مجموع المعاملات لنفس الثابت فى المعادلات من رقم 2 وحتى رقم $(r+1)$، وبنفس الطريقة فمعامل المعادلة رقم 1 يساوى مجموع المعاملات للمعادلات من رقم $(r+2)$ إلى $(r+c+1)$، ونفس الشىء ينطبق على الطرف الأيمن من المعادلات. وبالتالى لا يمكن إيجاد حلول لهذه المعادلات الآنية نظراً لوجود هذه العلاقة فيما بينها أى أنها تصبح معادلات غير مستقلة.

وحتى يمكن إيجاد حلول لهذه المعادلات الآنية (الاعتيادية) تطبق الاشتراطات السابق ذكرها فى (١٣-١٦) أو بعض الاشتراطات الأخرى والتى بتطبيقها يقل عدد المعادلات فى كل تقسيم classification معادلة واحدة، وكذلك عدد الثوابت المراد تقديرها، وبالتالى يصبح هناك معادلة واحدة للمتوسط μ، $(r-1)$ معادلات للمستويات الخاصة بالعامل α، $(c-1)$ معادلات للمستويات الخاصة بالعامل β أى يكون العدد الكلى للمعادلات الاعتيادية هو $(r+c-1)$.

فمثلاً إذا كان عدد المستويات للعامل α هو $r=3$ مستويات هو $\alpha_1,\alpha_2,\alpha_3$ فى حين أن عدد المستويات للعامل β هو مستويين $c=2$ وهما β_1,β_2 فإن العدد الكلى للمعادلات الاعتيادية الناتجة وهو 6 معادلات (واحدة للمتوسط + 3 للعامل α + 2 للعامل β)، أما بعد تطبيق الاشتراطات فيصبح العدد الكلى للمعادلات هو 4 (معادلة المتوسط + 2 للعامل α + معادلة واحدة للعامل β).

وعموماً يطلق على المعادلات الناتجة من تطبيق هذه الاشتراطات بالمعادلات الاعتيادية المختزلة reduced normal equation. وكنتيجة لحل هذه المعادلات المختزلة آنياً يمكن الحصول على التقديرات (الحلول) للثوابت التى تحقق أقل مربعات خطأ ممكن.

فإذا كان عدد المعادلات المطلوب حلها عدداً قليلاً فى حدود 3 أو 4 معادلات فإنه يمكن حلها بالطرق العادية لحل المعادلات الخطية، أما إذا زاد العدد عن ذلك، فإن جبر المصفوفات يصبح أنسب الطرق للحصول على تلك الحلول (الآن يجرى هذا بالحاسب الآلى فى أقل من لمح البصر).

ويلاحظ أيضاً فى جدول ١٣-٦ أن المعاملات للتقديرات المختلفة فى الجدول أى a_i، b_j تكون متناظرة حول الخط المحورى الرئيسى leading diagonal والذى يمثل القيم التى تتساوى فيها قيمة i, j (ويبدأ من الصف الأول والعمود الأول ثم الصف الثانى والعمود الثانى ... وهكذا حتى الصف الأخير والعمود الأخير). فمثلاً

جدول ١٣-٦ مصفوفة المعادلات الاعتيادية

المعلم parameter	المعادلة الاعتيادية Normal equation — الطرف الأيسر (LHS) $\hat{\mu}$	a_1	a_2	$\cdots$	a_r	b_1	b_2	$\cdots$	b_c	الطرف الأيمن (RHS)	رقم المعادلة
μ	$n_{..}$	$n_{1.}$	$n_{2.}$	$\cdots$	$n_{r.}$	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.c}$	$Y_{...}$	1
α_1	$n_{1.}$	$n_{1.}$				n_{11}	n_{12}	$\cdots$	n_{1c}	$Y_{1..}$	2
α_2	$n_{2.}$		$n_{2.}$			n_{21}	n_{22}	$\cdots$	n_{2c}	$Y_{2..}$	3
$\vdots$	$\vdots$			$\vdots$		$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
α_r	$n_{r.}$				$n_{r.}$	n_{r1}	n_{r2}	$\cdots$	n_{rc}	$Y_{r..}$	$r+1$
β_1	$n_{.1}$	n_{11}	n_{21}	$\cdots$	n_{r1}	$n_{.1}$				$Y_{.1.}$	$r+2$
β_2	$n_{.2}$	n_{12}	n_{22}	$\cdots$	n_{r2}		$n_{.2}$			$Y_{.2.}$	$r+3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$			$\vdots$		$\vdots$	$\vdots$
β_c	$n_{.c}$	n_{1c}	n_{2c}	$\cdots$	n_{rc}				$n_{.c}$	$Y_{.c.}$	$r+c+1$

المعامل للصف الأول والعمود الثانى $j = 2, i = 1$ هو n_{12} وهى نفس القيمة للصف الثانى والعمود الأول $j = 1, i = 2$ ... وهكذا. ويسمى هذا الترتيب للمعاملات فى الجدول بالمصفوفة المتناظرة symmetrical matrix.

وتعرف المصفوفة السابقة الذكر بمصفوفة التباين والتغاير للمعاملات
Variance-covariance coefficient matrix

ويلاحظ فى هذا الجدول أنه داخل كل تقسيم من التقسيمات تكون المعاملات الخارجة عن الخط المحورى الرئيسى off-diagonal تساوى صفراً. فمثلاً معاملات $a_2, a_3, \cdots, a_r$ فى معادلة α_1 تساوى صفراً. وأيضاً معاملات $b_2, b_3, \cdots, b_c$ فى معادلة β_2 تساوى صفراً أيضاً.

أما RHS فهى اختصار لـ Right Hand Side أى الجانب الأيمن وهى ببساطة تمثل مجاميع الأقسام المقابلة. فالمقابل لمعادلة المتوسط هو المجموع العام بينما $Y_{1..}$ هو مجموع المعاملة الأولى ... وهكذا.

وكما سبق فإنه بحل المعادلات الاعتيادية المختزلة يمكن الحصول على التقديرات للثوابت فى النموذج وهى قيم $\hat{\mu}$، a_i، b_j حيث تمثل هذه قيم الثوابت μ، α_i، β_j على التوالى.

ويمكن إثبات أن الاختزال reduction فى مجموع المربعات الكلى الراجع لموائمة النموذج الرياضى الكامل (١٣-١٥) يحسب كالتالى:

الاختزال فى مجموع المربعات الكلى نتيجة لموائمة جميع التقديرات μ، α_i، β_j
Reduction in total sum of squares due to fitting all constants
ويرمز له $R(\mu, a_i, b_j)$ وهو يساوى:

$$R(\mu, a_i, b_j) = \hat{\mu}Y_{...} + a_1 Y_{1..} + a_2 Y_{2..} + \cdots + a_r Y_{r..} + b_1 Y_{.1.} + b_2 Y_{.2.} + \cdots + b_c Y_{.c.} \qquad (١٣-١٨)$$

والتى يمكن اختصارها كالتالى:

$$R(\mu, a_i, b_j) = \hat{\mu}Y_{...} + \sum_i a_i Y_{i..} + \sum_j b_j Y_{.j.} \qquad (١٣-١٩)$$

ويعتبر مجموع مربعات الخطأ أو المتبقى المتحصل عليه من مواءمة هذه التقديرات للثوابت فى النموذج الرياضى هو أقل مجموع مربعات ممكن، نظراً لأنه قد استخدمت جميع الثوابت.

وبتطبيق نفس الطريقة العامة يمكن مواءمة نماذج مختزلة لأى مشاهدة تتضمن تأثير عامل واحد من العاملين والنماذج التى تطبق تكون كالتالى:

$$Y_{ijk} = \mu + \alpha_i + \xi_{ijk} \qquad (٢٠-١٣)$$

$$Y_{ijk} = \mu + \beta_j + \xi_{ijk} \qquad (٢١-١٣)$$

وبالتالى يمكن حساب الاختزال reduction فى مجموع المربعات نتيجة لموائمة النموذجين (١٣-٢٠)، (١٣-٢١). ويعبر عنها كالتالى $R(\mu,\alpha_i)$ ، $R(\mu,\beta_j)$ على التوالى.

ويلاحظ أنه بمواءمة النموذج الرياضى

$$Y_{ijk} = \mu + \xi_{ijk} \qquad (٢٢-١٣)$$

يحسب الاختزال فى مجموع المربعات نتيجة لهذا العامل، ويعبر عنه $R(\mu)$ وهذه الكمية تساوى

$$R(\mu) = \frac{(\Sigma Y_{ijk})^2}{n..} = CF \qquad (٢٣-١٣)$$

أى أنها تساوى قيمة معامل التصحيح، وعلى هذا الأساس فإنه يمكن باستخدام المعادلتين (١٣-١٩)، (١٣-٢٣) حساب الاختزال فى مجموع المربعات كنتيجة لمواءمة تأثير كل من العاملين A, B ويمكن التعبير عنها كالتالى:

$$R(a_i, b_j) = \hat{\mu}Y... + \sum_i a_i Y_i.. + \sum_j b_j Y._j. - CF \qquad (٢٤-١٣)$$

والآن يمكن استخلاص الاختزال الإضافى فى مجموع المربعات الراجع لمواءمة α_i والتى يرمز لها بالرمز $R'(a_i)$ وذلك عن طريق طرح معامل التصحيح من الاختزال فى مجموع المربعات الناتج من مواءمة النموذج $\mu+\alpha_i$ فى (١٣-٢٠)، وهو بالضبط يساوى ما تم الحصول عليه فى التحليل المبدئى أى مجموع المربعات الخاص بتأثير العامل A بإهمال أى تأثير للعامل B حسب المعادلة (١٣-١٢) وطبعاً نفس الشئ ينطبق على $R'(b_j)$. ومن جميع ما تقدم يمكن الآن، حساب تحليل التباين النهائى ووضعه فى جدول يمكن منه اختبار فرض العدم الخاص بالتأثيرات الرئيسية للعوامل A, B بناء على النموذج الرياضى الخالى من تأثير التداخل بين العوامل (١٣-١٥). ولقد سبق إيضاح أنه يمكن حساب مجموع المربعات الخاص بالتداخل بين

العاملين A، B من الحسابات الخاصة بهذا النموذج الرياضى بطريقة غير مباشرة ولو أنه غير وارد كجزء من التأثيرات، وبالتالى فإنه يمكن اختبار متوسط المربعات للتداخل لاختبار مدى تعبير النموذج (١٣-١٥) عن البيانات فى أحقية عدم وجود تداخل بين التأثيرات.

ومن المفترض أنه حتى يكون هناك تداخل بين العوامل وبالتالى يمكن حسابه مباشرة أن يتم تبنى النموذج الرياضى الخاص بوجود هذا التأثير، كما سبق إيضاحه، فى (١٣-٨) وبالاشتراطات السابق ذكرها فى (١٣-٦) أى أنه يتم مواءمة النموذج والحصول على الاختزال الكلى فى مجموع المربعات أى $R[\mu, a_i, b_j, (ab)_{ij}]$ وبالتالى $R[a_i, b_j, (ab)_{ij}]$ بطرح معامل التصحيح.

ولكن يلاحظ أن هذا الأخير هو بالضبط ما سبق الحصول عليه فى (١٣-١١) أى مجموع المربعات بين الفئات subclasses SS أى أنه يمكن الحصول عليه بالطرح. وعلى هذا الأساس يمكن عمل جدول تحليل التباين النهائى للبيانات كما سيوضح فى جدول ١٣-٧.

ومن هذا الجدول يمكن اختبار كل من التأثيرات الرئيسيـــة للنموذج (١٣-١٥) كما يمكن أيضاً اختبار معنوية التداخل بين العوامل، وبالطبع فإن اختبار التأثيرات يصبح صحيحاً فقط فى حالة عدم معنوية التداخل فى العشيرة.

ويتم اختبار معنوية التداخل بين العوامل حسب فرض العدم: $H_o : \alpha\beta = 0$، بقسمة متوسط مربعات التداخل على متوسط مربعات الخطأ أى:

$$F = (MS)_{AB} / (MS)_E$$

وذلك لكل أنواع التأثيرات سواء كانت ثابتة أو عشوائية أو مختلطة كما فى جدول ١٣-٧، وتكون درجات الحرية للاختبار هى درجات الحرية المقابلة لكل متوسط مربعات.

أما لاختبار كل من التأثيرات الرئيسية A، B فإنها أيضاً يتم اختبارها باستخدام متوسط المربعات للخطأ $(MS)_E$ حيث أنه فى هذه الطريقة هناك افتراض مبدئى أنه لا يوجد تداخل بين تأثير العوامل فى العشيرة، وبذا يصبح متوسط المربعات للخطأ هو الأساس للاختبار سواء لفرض العدم $H_o : \alpha = 0$ أو $H_o : \beta = 0$ وبدرجات الحرية المقابلة لكل أى $(r-1)$ أو $(c-1)$ للبسط على التوالى ودرجـــات حـــريـــة $(n..-rc)$ للمقام.

وهنا قد تجدر الملاحظة بأنه فى حالة عدم واقعية افتراض عدم وجود تداخل بين العوامل يصبح اختبار فرض العدم الخاص بالتأثيرات الرئيسية غير صحيح تماماً،

جدول ١٣-٧ جدول كامل لبيانات محللة بطريقة مواءمة الثوابت*

SOV	df	SS	MS
B باستبعاد A A eliminating B	r 1	R(a,b) SSB ignoring A R(a) (١٣-١٣) - (٢٤-١٣)	R(a)/(r 1) $(MB)_A$
A باستبعاد B B eliminating A	c 1	R(a,b) SSA ignoring B R(b) (١٢-١٣) - (٢٤-١٣)	R(b)/(c 1) $(MB)_B$
التداخل بين A، B AB interaction	(r 1)(c 1)	R[a,b,(ab)] R(a,b) R(ab) (٢٤-١٣) - (١١-١٣)	R(ab)/[(r 1)(c 1)] $(MB)_{AB}$
داخل الفئات Within subclasses	n.. rc	Total SS subclasses SS SEE (١٤-١٣)	SSE/(n.. rc) $(MS)_E$

* الأرقام تشير إلى أرقام المعادلات التى يمكن استخدامها فى الحسابات

ويلزم إتباع طرق أخرى لتحليل البيانات لا تتضمن أصلاً هذا الافتراض الخاص بالتداخل فى العشيرة مثل طريقة المربعات الموزونة للمتوسطات method of weighted squares of means.

مثال ١٣-٣

البيانات التالية فى جدول ١٣-٨ تمثل كيفية إجراء تحليل التباين باستخدام طريقة مواءمة الثوابت فى التقسيم الثنائى الجهة. وتمثل البيانات أوزان الفطام للحملان من 4 سلالات ومقسمة حسب جنس الحمل.

جدول ١٣-٨ وزن الفطام للحملان من 4 سلالات مقسمة حسب الجنس

الجنس		الإناث b_1	الذكور b_2	المجموع
العواسى a_1	n_{ij}	24	20	**44**
	$\sum Y_{ijk}$	648.0	603.0	**1251**
	$\sum Y_{ijk}^2$	17619.5	18688.0	**36307.5**
النجدى a_2	n_{ij}	19	22	**41**
	$\sum Y_{ijk}$	563.0	579.0	**1142**
	$\sum Y_{ijk}^2$	17155.5	15976.5	**33132**
الدوربر a_3	n_{ij}	19	13	**32**
	$\sum Y_{ijk}$	577.0	362.5	**939.5**
	$\sum Y_{ijk}^2$	17888.0	10483.75	**28371.75**
الحرى a_4	n_{ij}	12	14	**26**
	$\sum Y_{ijk}$	200.5	254.0	**454.5**
	$\sum Y_{ijk}^2$	3459.75	4636.5	**8096.25**
المجموع		**74**	**69**	**143**
		1988.5	**1798.5**	**3787**
		56122.75	**49784.75**	**105907.5**

ومن هذا الجدول يمكن استنباط المعادلات الاعتيادية الخاصة بهذا المثال كما هو موضح فى جدول ١٣-٦ وهذه المعادلات حسب النموذج الرياضى

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \xi_{ijk}$$

وكما سبق بافتراض عدم وجود تداخل بين العوامل ستكون المعادلات الاعتيادية كالتالى:

المتوسط $143\hat{\mu} + 44a_1 + 41a_2 + 32a_3 + 26a_4 + 74b_1 + 69b_2 = 3787$

العواسى (a_1) $44\hat{\mu} + 44a_1 + 24b_1 + 20b_2 = 1251$

النجدى (a_2) $41\hat{\mu} + 41a_2 + 19b_1 + 22b_2 = 1142$

الدوربر (a_3) $32\hat{\mu} + 32a_3 + 19b_1 + 13b_2 = 939.5$

الحرى (a_4) $26\hat{\mu} + 26a_4 + 12b_1 + 14b_2 = 454.5$

الإناث (b_1) $74\hat{\mu} + 24a_1 + 19a_2 + 19a_3 + 12a_4 + 74b_1 = 1988.5$

الذكور (b_2) $69\hat{\mu} + 20a_1 + 22a_2 + 13a_3 + 14a_4 + 69b_2 = 1798.5$

وهناك العديد من الاشتراطات التى يمكن فرضها لجعل المعادلات الاعتيادية ذات حل منها أن $a_1 + a_2 + a_3 + a_4 = 0$ ، $b_1 + b_2 = 0$. أو اشتراط آخر هو افتراض أى من a_1 أو a_2 أو a_3 أو a_4 مساويا للصفر وكذلك أى من b_1 أو b_2 مساويا للصفر. وبالتالى يمكن الحصول على المعادلات الاعتيادية المختزلة reduced normal equations وذلك بطرح أحد المعادلات من كل تقسيم. مثلاً طرح معادلة a_4 من كل من a_1، a_2، a_3 وأيضا طرح معادلة b_2 من معادلة b_1. وعن طريق الطرح من الصفوف والأعمدة فى كل معادلة يمكن الحصول على المصفوفة التالية:

المتوسط	$\hat{\mu}$	a_1	a_2	a_3	a_4	b_1	b_2	
μ	143	44	41	32	26	74	69	= 3787
α_1	18	44	0	0	-26	12	6	= 796.5
α_2	15	0	41	0	-26	7	8	= 687.5
α_3	6	0	0	32	-26	7	-1	= 485
β_1	5	4	-3	6	-2	74	-69	= 190

يلى ذلك طرح العمود a_4 من كل من العمود a_1 و a_2 و a_3 وأيضاً طرح العمود b_2 من العمود b_1، للحصول على المصفوفة التالية:

$$
\begin{array}{cc}
\begin{array}{ccccc} \hat{\mu} & a_1 & a_2 & a_3 & b_1 \end{array} & \text{RHS} \\
\begin{bmatrix} 143 & 18 & 15 & 6 & 5 \\ 18 & 70 & 26 & 26 & 6 \\ 15 & 26 & 67 & 26 & -1 \\ 6 & 26 & 26 & 58 & 8 \\ 5 & 6 & -1 & 8 & 143 \end{bmatrix} & \begin{bmatrix} 3787.0 \\ 796.5 \\ 687.5 \\ 485.0 \\ 190.0 \end{bmatrix} \\
X'X & X'Y
\end{array}
\quad
\begin{array}{l} \text{المتوسط } (\mu) \\ \text{العواسى } (\alpha_1) \\ \text{النجدى } (\alpha_2) \\ \text{الدوربر } (\alpha_3) \\ \text{الإناث } (\beta_1) \end{array}
$$

ويلاحظ فى المصفوفة الأخيرة والتى يعبر عنها بـــ $X'X$ ما يلى:

١- أن المصفوفة متناظرة حول القطر الرئيسى leading diagonal أى أن قيمة العنصر الموجود فى الصف الثانى والعمود الثالث (26) هى نفس قيمة العنصر الموجود فى الصف الثالث والعمود الثانى (26) ... وهكذا.

٢- القيم الموجودة على الخط القطرى الرئيسى لابد وأن تكون موجبة، وذلك لكونها تمثل مجموع مربعات أما القيم الموجودة خارج هذا القطر off-diagonal فيمكن أن تكون موجبة أو سالبة، وذلك لأنها تمثل مجموع حاصل ضرب cross products.

٣- أن قيمة العنصر فى الصف الأول والعمود الأول وهى معامل المتوسط $\hat{\mu}$ تمثل العدد الكلى للأفراد، وكذلك قيمة الجانب الأيمن (RHS) والذى يعبر عنه بـــ $X'Y$ لهذه المعادلة تمثل المجموع الكلى للقيم.

٤- تبعاً لتطبيق الاشتراطات لإيجاد الحلول للمعادلات فإن العلاقة الموجودة بين معادلة المتوسط $\hat{\mu}$ والمعادلات الخاصة بكل تأثير والتى كانت متمثلة فى أن معامل الثوابت فى معادلة المتوسط يساوى مجموع معاملات الثوابت فى المعادلة الخاصة بكل تأثير، وكذلك RHS قد اختفت وبالتالى أصبحت المعادلات مستقلة عن بعضها أى ذات حل.

وهناك أكثر من طريقة لحل المعادلات الاعتيادية المختزلة آنياً simultaneous solving منها حل المعادلات عن طريق جبر المصفوفات matrix algebra وبالتالى يلزم الحصول على مقلوب المصفوفة matrix inverse. وعلى هذا الأساس فإن مقلوب المصفوفة المختزلة $X'X$ وهو $(X'X)^{-1}$ يساوى:

$$\begin{bmatrix} 0.007309 & -0.001607 & -0.001228 & 0.000546 & -0.000227 \\ -0.001607 & 0.018690 & -0.004521 & -0.006128 & -0.000417 \\ -0.001228 & -0.004521 & 0.019575 & -0.006724 & 0.000746 \\ 0.000546 & -0.006128 & -0.006724 & 0.023098 & -0.001101 \\ -0.000227 & -0.000417 & 0.000746 & -0.001101 & 0.007085 \end{bmatrix}$$

وللحصول على تقديرات للثوابت فإن $\hat{\beta} = (X'X)^{-1}X'Y$ حيث $\hat{\beta}$ هى المتجهة vector: $(\hat{\mu} \quad a_1 \quad a_2 \quad a_3 \quad b_1)$ وبالتالى يمكن الحصول على التقديرات التالية للثوابت فى النموذج الرياضى:

١- المتوسط العام: $\hat{\mu} = 25.777$ kg

٢- تأثير السلالات (عددها 3 فى المصفوفة المختزلة):

$a_1 = 2.643$، $a_2 = 2.086$ ، $a_3 = 3.557$ للعواسى والنجدى والدروبر على التوالى. وبتطبيق الاشتراطات (١٣-١٦) يمكن حساب قيمة a_4 حيث:

$$a_4 = -(a_1 + a_2 + a_3) = -8.286$$

وبالتالى يمكن حساب متوسط كل من السلالات بجمع التأثير الخاص بتلك السلالة على قيمة المتوسط العام للحصول على المتوسطات التالية:

العواسى $25.777 + 2.643 = 28.42$

النجدى $25.777 + 2.086 = 27.862$

الدوربر $25.777 + 3.557 = 29.334$

الحرى $25.777 + (-8.286) = 17.49$

٣- تأثير جنس الحمل (عددها واحد فى المصفوفة المختزلة) $b_1 = 0.132$ kg للإناث، وبتطبيق الاشتراط السابق فإن قيمة $b_2 = -0.132$، وعلى هذا الأساس يكون متوسط الإناث عند الفطام $25.777 + 0.132 = 25.909$ ومتوسط الذكور $25.777 + (-0.132) = 25.645$

ويمكن حالياً البدأ فى حساب مجاميع المربعات حتى يمكن استكمال تحليل التباين:

١- مجموع المربعات الكلية الغير مصحح: $\sum_{ijk} Y_{ijk}^2 = 105907.5$

٢- معامل التصحيح:

$$CF = \frac{(Y...)^2}{n} = \frac{(3787)^2}{143} = 100289.2937$$

٣- مجموع المربعات الكلي المصحح:

$$TSS = \sum_{ijk} Y_{ijk}^2 - CF = 105907.5 - 100289.2937 = 5618.2063$$

٤- مجموع المربعات بين الفئات:

$$\text{Subclasses SS} = \sum_{ij} \frac{Y_{ij}^2}{n_{ij}} - CF = \frac{(648)^2}{24} + \frac{(603)^2}{20} + \cdots + \frac{(254)^2}{14} - CF$$
$$= 2897.0211$$

٥- مجموع المربعات الخاص بالعامل الأول (a) بإهمال تأثير العامل الثانى (b):

$$\text{SSA ignoring B} = \sum_{i} \frac{Y_{i..}^2}{n_{i.}} - CF = \frac{(1251)^2}{44} + \cdots + \frac{(454.5)^2}{20} - CF$$
$$= 2615.9313$$

٦- مجموع المربعات للعامل الثانى (b) بإهمال تأثير العامل الأول (a):

$$\text{SSB ignoring A} = \sum_{j.} \frac{Y_{.j.}^2}{n_{.j}} - CF = \frac{(1798.5)^2}{69} + \frac{(1988.5)^2}{74} - CF = 23.2194$$

٧- مجموع المربعات داخل الفئات (الخطأ) Within subclasses SS:

$$\text{Within subclasses SS} = \sum_{ijk} Y_{ijk}^2 - \sum_{ij} \frac{Y_{ij.}^2}{n_{ij}} = 105907.5 - 103186.3148$$
$$= 2721.1852$$

ويستكمل تحليل التباين باستعمال طريقة مواءمة الثوابت واستخدام تقديرات الثوابت السابق الحصول عليها.

٨- يحسب الاختزال فى مجموع المربعات لمواءمة جميع الثوابت μ، a_i، b_j حسب المعادلة (١٣-١٩):

$$R(\mu, a_i, b_j) = \hat{\mu}\, Y_{...} + \sum_i \hat{a}_i \; Y_{i..} + \sum_j \hat{b}_j \; Y_{.j.} = \hat{\beta}(X'Y)$$

$$= \begin{bmatrix} 25.777 & 2.643 & 2.086 & 3.557 & 0.132 \end{bmatrix} \begin{bmatrix} 3787 \\ 796.5 \\ 687.5 \\ 485 \\ 190 \end{bmatrix}$$

$$= (25.77)(3787) + (2.643)(796.5) + \cdots + (0.132)(190) = 102907.6872$$

ويفضل فى الحسابات السابقة أن يتم استخدام قيم الثوابت بدون تقريب أى 6 أرقام عشرية على الأقل.

٩- يحسب الاختزال فى مجموع المربعات نتيجة لمواءمة العوامل غير المتوسط حسب المعادلة (١٣-٢٤):

$$R(a_i, b_j) = R(\mu, a_i, b_j) - CF = 102907.6872 - 100289.2937 = 2618.3935$$

١٠- مجموع المربعات للتداخل بين العاملين A، B:

$$\begin{aligned} \text{SS interaction (A x B)} &= R[a_i, b_j,] - R(a_i, b_j) \\ &= 2897.0211 - 2618.3935 = 278.6276 \end{aligned}$$

١١- مجموع المربعات للعامل A باستبعاد B أى $R(a_i)$:

$$\begin{aligned} \text{SSA eliminating B} &= R(a_i) = R(a_i, b_j) \text{ - SSB ignoring A} \\ &= 2618.3935 - 23.2194 = 2595.1741 \end{aligned}$$

١٢- مجموع المربعات للعامل B باستبعاد A أى $R(b_j)$:

$$\begin{aligned} \text{SSB eliminating A} &= R(b_j) = R(a_i, b_j) \text{ - SSA ignoring B} \\ &= 2618.3935 - 23.2194 = 2595.1741 \end{aligned}$$

ويمكن تلخيص النتائج فى جدول ١٣-٩ لتحليل التباين كالتالى:

جدول ١٣-٩ تحليل التباين النهائى لبيانات مثال ١٣-٨

SOV	df	SS	MS	F
Factor A, $R(a_i)$	$c-1=3$	2595.174	865.058^{**}	42.9
Factor B, $R(b_j)$	$r-1=1$	2.462	2.462	< 1
Interaction A x B	$(r-1)(c-1)=3$	278.628	92.88^{**}	4.6
Error	$n_{..}-rc=135$	2721.185	20.16	
Total	$n_{..}-1=142$			

ويتضح من نتيجة اختبار F أنه ولو أن النموذج المفترض كان لا يحتوى على التأثير الخاص بالتداخل بين العاملين A، B إلا أن تأثيره كان معنوياً عند مستوى (P<0.01) وبالتالى فإنه كان من الأوفق عند افتراض وجود هذا التأثير فى النموذج مسبقاً أن تستخدم طريقة المربعات الموزونة للمتوسطات (الفصل ١٣-٧) method of weighted squares of means. أو أنه كنتيجة لوجود تداخل معنوى بين العوامل فإنه لا يصح اختبار التأثيرات الرئيسية للعوامل main effects بل أنه يجب استخدام طريقة المربعات الموزونة للمتوسطات لاستكمال تحليل التباين. ومن المهم أن يلاحظ أنه إذا كانت البيانات تحتوى على بعض الفئات subclasses الخالية من الأفراد أو الفارغة empty cells فإنه لا يمكن استخدام غير طريقة مواءمة الثوابت السابق ذكرها حيث أن الطرق الأخرى تحتاج إلى حساب الكمية $1/n_{ij}$ وهو ما لا يمكن استخدامه فى تلك الحالات.

وهناك الكثير من الإحصائيين ينصحون بأنه فى حالة استخدام طريقة مواءمة الثوابت تختبر معنوية التداخل بين العوامل ثم بناء على نتيجة هذا الاختبار يتم اختبار التأثيرات الرئيسية. وهو أن يتم اختبار متوسط مربعات التداخل على مستوى معنوية منخفض أى حوالى (P<0.25) حتى لا يحدث تأثيراً قوياً على مستوى معنوية الاختبارات اللاحقة للتأثيرات الرئيسية.

١٣-٧ طريقة المربعات الموزونة للمتوسطات

Method of weighted squares of means

كما ذكر سابقاً، فإنه إذا وجد أنه باختبار متوسط المربعات للتداخل بين العوامل كان معنوياً فإنه لا يصبح اختبار التأثيرات الرئيسية غير منحاز. وطريقة المربعات الموزونة للمتوسطات تقدم اختباراً غير منحاز للتأثيرات الرئيسية فى حالة معنوية

التداخل، ويتم الوزن حسب مقلوب تباين متوسطات الخلايا والذى يكون محصلة لعدد الأفراد فى كل خلية.

ويتم تحليل متوسطات الخلايا أو الفئات cell or subclass means حيث يعرف متوسط أى خلية

$$\bar{Y}_{ij.} = \sum_k Y_{ijk} / n_{ij} \qquad (١٣-٢٥)$$

والنموذج الرياضى المفترض فى هذه الحالة لكل قيمة يحتوى على التداخل بين التأثيرات الرئيسية أى $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \xi_{ijk}$ حيث $i = 1, 2, \cdots, r$، $j = 1, 2, \cdots, c$، $k = 1, 2, \cdots, n_{ij}$، وقد سبق الحديث عن الافتراضات الخاصة بهذا النموذج. وإتباع طريقة التحليل بالمربعات الموزونة للمتوسطات مبنيا على أساس ما أشير إليه سابقا فى أن تباين متوسط القيم يساوى تباين القيمة الواحدة مقسوما على عدد القيم أى أنه إذا افترض أن:

$$\bar{Y}_{ij.} = \left[\sum_j \left(\frac{Y_{ij\cdot}}{n_{ij}} \right) / c \right] \qquad (١٣-٢٦)$$

فإن التباين لهذا المتوسط

$$V(\bar{Y}_{i}..) = V\left[\sum_j \left(\frac{Y_{ij\cdot}}{n_{ij}} \right) / c \right] = \frac{1}{c^2} \left[\sum_j V\left(\frac{Y_{ij\cdot}}{n_{ij}} \right) \right] \qquad (١٣-٢٧)$$

وبافتراض أن التباينات للخلايا كلها متجانسة homogeneous وقيمة كل منها σ^2 فإن (١٣-٢٧) تصبح:

$$\frac{1}{c^2} \sum_j \frac{\sigma^2}{n_{ij}} = \frac{\sigma^2}{c^2} \sum_j \frac{1}{n_{ij}} = \frac{\sigma^2}{\omega_i} \qquad (١٣-٢٨)$$

حيث

$$\frac{1}{\omega_i} = \frac{1}{c^2} \sum_j \frac{1}{n_{ij}} \qquad (١٣-٢٩)$$

وتمثل c عدد الأعمدة أى مستويات العامل B، وتعرف القيمة ω_i (وتنطق Omega) على أنها الوزن الخاص بالمتوسط $\overline{Y}_{i..}$ وسوف يتم استخدام هذا الوزن لحساب مجموع المربعات للتأثير الرئيسى للعامل A.

وبنفس الطريقة توجد قيمة:

$$V(\overline{Y}_{.j.}) = \sigma^2 / \nu_j \qquad (١٣-٣٠)$$

حيث

$$\frac{1}{\nu_j} = \frac{1}{r^2} \sum_i \frac{1}{n_{ij}} \qquad (١٣-٣١)$$

وتمثل r عدد الصفوف أو مستويات العامل A، وتكون ν_j (وتنطق Nu) هى الوزن الخاص بالمتوسط $\overline{Y}_{.j.}$ وتستخدم فى حساب مجموع المربعات للتأثير الرئيسى للعامل B.

وبناء على ذلك فإنه للحصول على مجموع المربعات الخاص بالتأثيرات الرئيسية بالنسبة للعامل A فإن مجموع المربعات الموزون الخاص به يمكن حسابه كالتالى:

$$SSA = \sum_i \omega_i \overline{Y}_{i..}^2 - \frac{(\sum_i \omega_i \overline{Y}_{i..})^2}{\sum_i \omega_i} \qquad (١٣-٣٢)$$

أما مجموع المربعات الموزون للعامل B فيكون كالتالى:

$$SSB = \sum_j \nu_j \overline{Y}_{.j.}^2 - \frac{(\sum_j \nu_j \overline{Y}_{.j.})^2}{\sum_j \nu_j} \qquad (١٣-٣٣)$$

ويلاحظ أن عملية الوزن باستخدام مقلوب الأعداد لكل فئة تؤدى إلى أن المشاهدات التى يكون تباينها أقل (نظراً لكثرة العدد فى الفئة) تكون الدقة فى تقديرها أكبر وبالتالى فإنه يلزم إعطاؤها وزناً أكبر عند حساب مجموع المربعات.

ولاستكمال تحليل التباين فإن التحليل يبدأ باستخدام طريقة مواءمة الثوابت المذكورة سابقاً، ويحسب من هذا مجموع المربعات بين الفئات between subclasses sum of squares كما فى (١٣-١١) ثم يحسب الاختزال فى مجموع المربعات كنتيجة لمواءمة تأثيرى كل من العاملين الرئيسين A، B كما فى (١٣-٢٤) وبالتالى يمكن الحصول على مجموع المربعات للتداخل بطرح الثانى من الأول كما سبق. بعد ذلك يتم اختبار متوسط المربعات للتداخل، فإن وجد أنه معنوى لا يستكمل التحليل بطريقة مواءمة الثوابت بل يتم الحصول على مجاميع المربعات للتأثيرات الرئيسية بالطريقة التى شرحت فى هذا الجزء. هذا مع العلم بأنه فى أى من الطرق لابد من حساب مجموع المربعات داخل الفئات within-subclasses sum of squares أولا كما فى (١٣-١٤) حتى يمكن اختبار معنوية التداخل باستخدام متوسط المربعات داخل الفئات، ويستكمل تحليل التباين بعد ذلك بالطريقة العادية لاختبار متوسط المربعات لكل من التأثيرين الرئيسين.

صندوق ١٣-٢

طريقة موائمة الثوابت وطريقة المربعات الصغرى الموزونة: هما طريقتان مضبوطتان لحساب تحليل التباين، الأولى عندما يكون التداخل بين العوامل الرئيسية غير معنوى والثانية عندما يكون هذا التداخل معنوياً.

والفكرة فيهما أنهما يقدران المعالم بما يحقق أن الخطأ فى النموذج هو أقل ما يمكن لذا أطلق عليهما "أقل مربعات للخطأ" وإن كانت الطريقتان كثيرة المعادلات والحساب ولكن أساسهما الرياضى بسيط ويجرى التحليل بواسطة الحواسب فى ومضة عين.

مثال ١٣-٤

باستخدام البيانات الموجودة فى جدول ١٣-٨ والخاصة بأوزان الحملان عند الفطام يتضح من تحليل البيانات باستخدام طريقة مواءمة الثوابت (١٣-١٥) أن مجموع المربعات للتداخل المحسوب كان 278.6276 وله 3 درجات حرية، وبالتالى فإن متوسط المربعات للتداخل كان 92.88 وهى قيمة معنوية على مستوى $(P<0.01)$ كما هو واضح من جدول ١٣-٩، وعليه فإن هذا يستوجب عدم استكمال تحليل البيانات بهذه الطريقة لاختبار التأثيرات الرئيسية للعوامل. ويجب استخدام طريقة المربعات الموزونة للمتوسطات، ويوضح جدول ١٣-١٠ الحسابات المطلوبة لاستكمال تحليل البيانات.

وبناء على نتائج جدول ١٣-١٠ فإنه يمكن حساب مجموع المربعات للتأثيرات الرئيسية للعاملين على النحو التالى:

١- باستخدام المعادلة (١٣-٣٢).

$$\text{Breed SS} = 101587.887 - \frac{(3737.468)^2}{141.14} = 2617.595$$

٢- بتطبيق المعادلة (١٣-٣٣)

$$\text{Sex SS} = 89800.014 - \frac{(3483.224)^2}{135.11} = 3.083$$

ويمكن تلخيص نتائج تحليل التباين فى الجدول التالى:

SOV	df	SS	MS	F
Breed (B)	3	2617.595	872.532	43.28**
Sex (S)	1	3.083	3.083	< 1
B x S	3	278.628	92.876	4.61**
Within	135	2721.185	20.157	
Total	**142**			

جدول ١٣-١٠ الحسابات المطلوبة لاستكمال تحليل التباين للبيانات التى فى جدول ١٣-٨

السلالة		إناث	ذكور	$\sum_j(1/n_{ij})$	ω_i	$\overline{Y}_{i..}$	$(\omega_i)(\overline{Y}_{i..})$	$(\omega_i)(\overline{Y}_{i..})^2$
	n_{ij}	24	20					
العواسى	$1/n_{ij}$	0.04167	0.05	0.09167	43.63478	28.575	1246.8638	35629.133
	$\overline{Y}_{ij.}$	27.0	30.15					
	n_{ij}	19	22					
النجدى	$1/n_{ij}$	0.05263	0.04545	0.09808	40.78303	27.9749	1140.9011	35629.133
	$\overline{Y}_{ij.}$	29.6316	26.31818					
	n_{ij}	19	13					
الدوربر	$1/n_{ij}$	0.05263	0.07692	0.12955	30.87611	29.1265	899.3130	26193.84
	$\overline{Y}_{ij.}$	30.3684	27.8846					
	n_{ij}	12	14					
الحرى	$1/n_{ij}$	0.08333	0.07143	0.15476	25.84647	17.4256	450.3902	7848.3201
	$\overline{Y}_{ij.}$	16.7083	18.1429					
$\sum_j(1/n_{ij})$		0.23026	0.2438	المجموع	141.14039		3737.468	101587.887
v_j		69.4867	65.6276	135.1143				
$\overline{Y}_{.j.}$		25.9271	25.6239					
$(v_j)(\overline{Y}_{.j.})$		1801.5886	1681.635	3483.224				
$(v_j)(\overline{Y}_{.j.})^2$		46709.967	43090.047	89800.014				

مثال ١٣-٥

استخدام برنامج SAS لحل بيانات الجدول التالى والتى تمثل عدد الأيام اللازمة لإنبات germination ثلاثة أنواع varieties من بذور نبات معين فى نوعين من التربة soil باستخدام عدد مختلف من صوانى الإنبات pots (Searle, 1987).

نوع التربة	نوع بذور النبات ١	٢	٣
١	6	13	14
	10	15	22
	11		
٢	12	31	18
	15		9
	19		12
	18		

```
DATA GERMIN;
INPUT VARIETY SOIL DAYS @@;
CARDS;
1 1 6 1 1 10 1 1 11 2 1 13 2 1 15 3 1 14 3 1 22
1 2 12 1 2 15 1 2 19 1 2 18 2 2 31 3 2 18 3 2 9 3 2 12
PROC GLM;
CLASS VARIETY SOIL;
MODEL DAYS = VARIETY SOIL VARIETY*SOIL/SS3;
LSMEANS VARIETY SOIL VARIETY*SOIL/STDERR;
RUN;
```

لاحظ:

لابد من استخدام SS3 (النوع الثالث لمجموع المربعات) فى حالة عدم تساوى التكرارات داخل الفئات والذى يمثل الحل المضبوط بطريقة الاختزال reduction.

استخدام اختيار LSMEANS للحصول على متوسطات أقل مجموع مربعات والتى تكون أدق من المتوسط الحسابى وهى تعنى حساب القيمة المتوقعة لمتوسطات الفئات كما لو كانت البيانات متزنة أخذا فى الاعتبار القيمة المتوسطة للعوامل الرئيسية. استخدام اختيار STDERR للحصول على الخطأ المعيارى لهذه المتوسطات.

النتائج التحليل

The GLM Procedure

Class Level Information

Class	Levels	Values
VARIETY	3	1 2 3
SOIL	2	1 2

Number of observations 15

Dependent Variable: DAYS

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	400.0000000	80.0000000	6.00	0.0103
Error	9	120.0000000	13.3333333		
Corrected Total	14	520.0000000			

R-Square	Coeff Var	Root MSE	DAYS Mean
0.769231	24.34322	3.651484	15.00000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
VARIETY	2	192.1276596	96.0638298	7.20	0.0135
SOIL	1	123.7714286	123.7714286	9.28	0.0139
VARIETY*SOIL	2	222.7659574	111.3829787	8.35	0.0089

<u>Least Squares Means</u>

VARIETY	DAYS LSMEAN	Standard Error	Pr > |t|
1	12.5000000	1.3944334	<.0001
2	22.5000000	2.2360680	<.0001
3	15.5000000	1.6666667	<.0001

SOIL	DAYS LSMEAN	Standard Error	Pr > |t|
1	13.6666667	1.4054567	<.0001
2	20.0000000	1.5315610	<.0001

VARIETY	SOIL	DAYS LSMEAN	Standard Error	Pr > |t|
1	1	9.0000000	2.1081851	0.0021
1	2	16.0000000	1.8257419	<.0001
2	1	14.0000000	2.5819889	0.0004
2	2	31.0000000	3.6514837	<.0001
3	1	18.0000000	2.5819889	<.0001
3	2	13.0000000	2.1081851	0.0002

٨-١٣ تنويه عن تحليل البيانات غير متناسبة التكرار

تبين فيما سبق مدى تعقيد تحليل البيانات غير المتزنة وإن كانت الحاسبات وبرمجياتها قد يسرت كثيرا هذه التحليلات. ولكن يلزم الحذر عند تفسير البيانات غير متناسبة التكرار من حيث ما إذا كان الاختلاف فى التكرار يعكس التكرار الفعلى فى العشيرة أم إنه نتيجة لأثر العوامل المختلفة محل الدراسة كأن تؤدى معاملة معينة إلى ارتفاع نسبة النفوق مثلا. ويستلزم الأمر كثيرا من الحذر إذا كانت هناك خلايا فارغة empty cells ليست بها مشاهدات وهذا يؤثر على تقدير التداخلات بين العوامل وقد يصل الأمر إلى التأثير أيضا على تقدير العوامل الرئيسية main effects مما قد يستوجب تجزئة البيانات إلى أجزاء يحلل كل منها مستقلا أو بطريقة ما بين التحليل العنقودى nested analysis والتحليل المتعامد crossed analysis، والمثال التالى يوضح هذا المفهوم.

مثال ١٣-٦

افترض وجود عاملان A، B وبكل منهما ٤ مستويات. وفى الجدول التالى الخلايا التى بها علامة x هى خلايا بها مشاهدات أما الخلايا التى لا يوجد بها هذه العلامة فهى خلايا فارغة لا يوجد بها أى مشاهدات.

	A_4	A_3	A_2	A_1		
جزء I	x	x			B_1	
جزء I	x	x			B_2	
			x	x	B_3	جزء II
			x	x	B_4	جزء II

مثل هذه البيانات لا يمكن تحليلها كوحدة واحدة حيث لا يمكن مقارنة A_1 أو A_2 مع A_3 أو A_4 لأن المقارنتين A_1، A_2 يقعان تحت مستويات مختلفة من

B (B_3 ، B_4) والمقارنتين A_3 ، A_4 يقعان تحت B_1 ، B_2. ولكن يمكن مقارنة A_1 مع A_2 وكذلك مقارنة A_3 مع A_4. وتسمى مثل هذه البيانات بالبيانات غير المتصلة disconnected data والتى تتبع ظاهرة تسمى "الاتصالية" connectedness.

ويتم تحليل مثل هذه البيانات بتقسيمها إلى جزئين وليكن جزء I ويشمل أربع خلايا هى A_3B_1 ، A_3B_2 ، A_4B_1 ، A_4B_2 وجزء II ويشمل أربع خلايا هى A_1B_3 ، A_1B_4 ، A_2B_3 ، A_2B_4. ويكون التحليل كما يلى:

الجزء الأول I

SOV	df
A	1
B	1
A x B	1
الكلى	3

الجزء الثانى II

SOV	df
A	1
B	1
A x B	1
الكلى	3

ويمكن ضم التحليلين معا ليصبحا

SOV	df
A داخل الأجزاء	2
B داخل الأجزاء	2
A x B داخل الأجزاء	2

وهذا ما أطلق عليه خليط بين التحليل العنقودى والتحليل المتعامد.

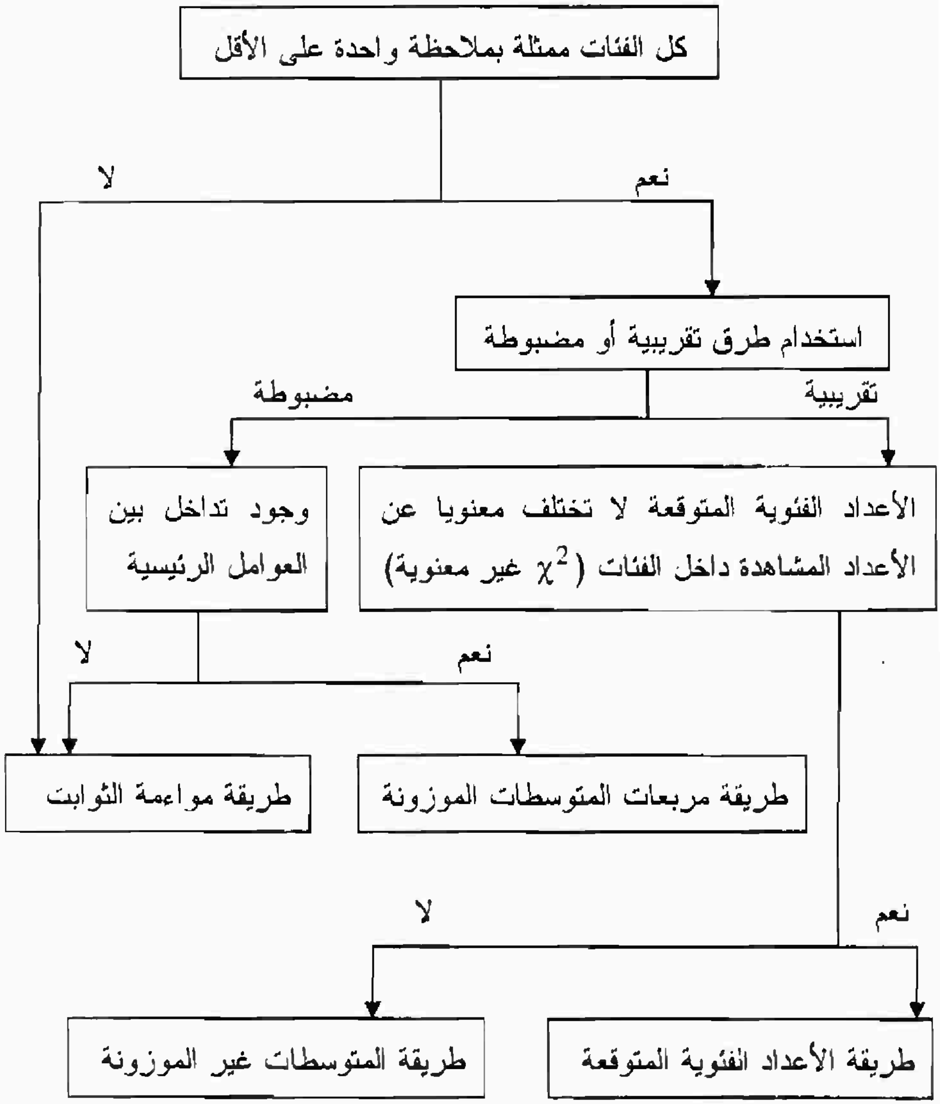
أما إذا فرض أن الخلية A_2B_2 أو A_3B_3 بأى منهما ولو فرد واحد تصبح البيانات متصلة connected data فإنه يجوز مقارنة كل العوامل السابقة ولكن ليست التداخلات كلها ويصبح التحليل:

SOV	df
A	3
B	3
A x B	2
الكلي	8

لاحظ أن درجات الحرية للتداخل ليست حاصل ضرب درجات حرية العوامل الرئيسية بسبب وجود الخلايا الفارغة.

أما إذا زاد الاتصال بوجود خلايا أخرى بها مشاهدات فإنه يمكن تقدير بعض التداخلات بدرجة أكبر ... وهكذا. وكلما زاد عدد الخلايا المحتوية على مشاهدات زادت الاتصالية وزاد التحليل رصانة وأمكن تقدير عدد أكبر من المعالم الإحصائية. ولمزيد من التفاصيل يمكن الرجوع إلى Searle, 1987.

ويبين شكل ١٣-١ هيكل توضيحى للطرق المستخدمة فى تحليل التباين فى حالة عدم تساوى تكرار الفئات.



شكل ١٣-١ هيكل توضيحى للطرق المستخدمة فى تحليل التباين فى حالة عدم تساوى تكرار الفئات

تمارين الباب الثالث عشر

١٣-١ البيانات المدرجة فى الجدول التالى تمثل متوسط وزن الذكور بالكيلوجرام عند عمر 8 أسابيع لأربع سلالات من الدجاج ناتجة من 3 دفعات فى أوقات مختلفة من السنة (لاحظ اختلاف عدد الطيور التى تصل إلى نهاية التجربة).

السلالة		دفعة الفقس		
		١	٢	٣
أ	الوزن	1.317	1.493	1.425
	العدد	18	11	11
ب	الوزن	1.641	1.675	1.471
	العدد	22	9	6
ج	الوزن	1.879	2.128	1.894
	العدد	17	8	9
د	الوزن	1.591	1.811	1.713
	العدد	26	7	18

فإذا كان مجموع المربعات الكلى المصحح 13.1525 . كون واملأ جدول تحليل التباين بطريقة المتوسطات غير الموزونة وقارنها بطريقة الأعداد المتوقعة للفئات.

١٣-٢ جربت ثلاث طرق محددة لتدريس الإحصاء على مجموعة من الطلبة والطالبات وفى نهاية التجربة كانت نتائج الاختبار كما يلى:

	طلبة	طالبات
طريقة ١	5، 9	7، 8، 3
طريقة ٢	4، 5، 7	6، 4، 9
طريقة ٣	2، 1، 4	

كون واملأ جدول تحليل التباين. اجر اختبارات المعنوية المختلفة عند مستوى معنوية 5% مع كتابة النموذج الإحصائى.