الفصل الحادي عشر

تحليل التشتت - التصميم التام العشوائية

١١ ـ ١ مقدمة : تضطرنا الظروف غالباً لتصميم تجربة ندرس فيها عدة مجتمعات في وقت واحد . واذا رغبنا في تقصي الفروق بين خمسة متوسطات ، أي مقارنة خمسة مجتمعات ، بالاعتماد على الاختبار **t** بالنسبة لكل زوج منها على حدة ، فلن تكون مثل هذه الطريقة الإحصائية جيدة لعدة أسباب ، أهمها أنه إذا استخدمنا مستوى من الأهمية يساوي% 5 عند كل إختبار على حدة فإن مستوى الأهمية بالنسبة لنتيجة نستقرؤها حول المتوسطات الخمسة معاً سيكون أكبر بكثير من % 5 . هذا بالإضافة إلى الخسارة في دقة تقدير التشتت نتيجة لإستخدامنا المعلومات المتوفرة في تلك القياسات المتعلقة بالمتوسطين اللذين نقارنهما فقط .

وتوضح الأمثلة التي سنقدمها في الفقرة القادمة أنواعاً من تصميمات التجارب نحللها وفق طرق تسمى طرق تحليل التشتت . ويعتمد التحليل ، كما سنرى في الفقرات القادمة ، على فصل تشتت كافة الملاحظات التي نحصل عليها من التجربة إلى أجزاء ، يشكل كل منها قياساً للتغير يعود إلى مصدر محدد من مصادر التغير ، فمثلاً تغيرات داخلية ضمن كل من المجتمعات المدروسة ، تغيرات من مجتمع إلى آخر ، الخ . وتشير عبارة تحليل التشتت إلى عملية تفكيك تشتت العينة إلى عدة مركبات . ويجب ألا ننسى أن المسألة الأساسية تتعلق بمقارنة متوسطات عدة مجتمعات ، وأننا

نقوم بتفكيك أو تحليل التشتت لهذه الغاية . وسنرى أنه توجد عدة طرق للقيام بمثل هذا التحليل ويعتمد ذلك على هدف التحليل فقد نرغب مثلاً القيام باختبار إجمالي حول تساوي أو عدم تساوي متوسطات المجتمعات المدروسة ، وقد نرغب في التعرف من بين هذه المجتمعات على المجتمع ذي المتوسط الأكبر

١١ ـ ٢ مناقشة أمثلة توضيحية لبعض المسائل :

مثال ١ ـ ١ التصنيف الأحادي : نجد أبسط تطبيقات طريقة تحليل التشتت في مسألة إختبار فرضية تتعلق بمتوسطات k من المجتمعات $\mu_1, \mu_2, \ldots, \mu_k$ أو تقدير هذه المتوسطات وتسمى مسألة التصنيف الأحادي . وقد ناقشنا في الفصل السابق الحالة التي تكون فيها k=2 ، حيث إختبرنا الفرضية $\mu_1 = \mu_2$ ،ووضعنا تقديراً مجالياً للفرق $\mu_1 - \mu_2$. ونصنف جميع الأفراد بحيث ينتمي كل منها إلى واحد من هذه المجتمعات فقط وينبغي التأكيد هنا على أن هذه المجتمعات الـ k تستنفذ كل الحالات التي نهتم بها ، وللتمييز بين هذه الحالة والحالة التي تكون فيها المجتمعات الـ k مجرد عينة من عدد أكبر من المجتمعات التي تتطرق إليها الدراسة (وسنتعرض لمثل هذه الحالة في المثالين ٤ و ٥) فسنشير إلى حالتنا هنا على أنها تمثل « نموذج الثوابت » أو « النموذج I » .

ونسوق فيما يلي أمثلة عن تصنيف أفراد مجتمع وفقاً لمتحول واحد :

(أ) لدينا خمس طرق للتعليم (k=5) ومجموعة من أطفال المدارس . ولنتصور خمسة مجتمعات افتراضية يحوي الواحد منها كل الأطفال الذين يتلقون تعليمهم بإحدى هذه الطرق . والمطلوب مقارنة متوسطات المعرفة المكتسبة (مقدرة بواسطة إختبار معين) في كل من الطرق الخمسة .

(ب) لدينا ثلاث مدن (k=3) ونريد مقارنة متوسط دخل السكان

في كل من هذه المدن . ونلاحظ في هذا المثال أن المجتمع هو سكان المدينة . بينما كان المجتمع في المثال السابق غير ملموس أو افتراضي .

(حـ) لدينا أربعة أصناف من القمح $(k=4)$ ونتصور مجتمع الانتاج لكل من هذه الأصناف . ونريد مقارنة متوسط إنتاج الأصناف الأربعة .

مثال ٢ ـ التصنيف الثنائي : نصنف الأفراد في هذا المثال وفقاً لخاصتين أو صفتين . وبصورة كيفية تدعى إحدى الصفتين بالمتحول الأول ، ولنفرض أن هذه الصفة تبدو في r من الأشكال ، كما تدعى الصفة الأخرى بالمتحول الثاني ، ولنفرض أنها تبدو في c من الأشكال . وقد وضعنا في الجدول (١١ ـ ١) المتحول الثاني في أعلى الجدول وسنسميه أحياناً متحول الأعمدة . كما نسمي المتحول الأول بمتحول الصفوف . وهكذا يتألف الجدول من c عموداً و r صفاً . ويمثل الجدول (١١ ـ ١) الحالة $c = 3$ و $r = 4$. ويمكن

جدول (١١ ـ ١) التصنيف الثنائي

المتحول الثاني

المتحول الأول	1	2	3	
1	μ_{11}	μ_{12}	μ_{13}	$\mu_{1.}$
2	μ_{21}	μ_{22}	μ_{23}	$\mu_{2.}$
3	μ_{31}	μ_{32}	μ_{33}	$\mu_{3.}$
4	μ_{41}	μ_{42}	μ_{43}	$\mu_{4.}$
	$\mu_{.1}$	$\mu_{.2}$	$\mu_{.3}$	$\mu_{..}$

إعتبار كل من الخلايا الاثنتي عشرة في الجدول كمجتمع خاص ، وينتمي كل فرد إلى واحد منها فقط . ونهتم بوجود أو عدم وجود فرق بين متوسطات هذه المجتمعات . ونرمز بـ μ_{ij} لمتوسط المجتمع الخاص بالخلية (ij) ،

حيث يشير الدليل الأول i إلى الصف الذي يحوي هذه الخلية ، ويرمز الدليل الثاني j إلى العمود الذي تقع فيه الخلية (ij) . و μ_{23} مثلاً هو متوسط مجتمع الأفراد الذين يجمعون الصفة 2 من المتحول الأول والصفة 3 من المتحول الثاني . وتظهر على الجانب الأيمن من الجدول مقادير من النوع $\mu_{i.}$ نعرفها بأنها متوسط الصف i ، ومثلاً $\mu_{1.}=(\mu_{11}+\mu_{12}+\mu_{13})/3$ ، وفي أسفل الجدول تظهر المقادير $\mu_{.j}$ نعرفها بأنها متوسط العمود j ، فمثلاً $\mu_{.1}=(\mu_{11}+\mu_{21}+\mu_{31}+\mu_{41})/4$. والمتوسط العام للمقادير μ_{ij} الاثني عشر نرمز له بـ $\mu_{..}$ وقد سجلناه في الزاوية السفلية اليمنى من الجدول .

ونفترض كما في المثال الأول أن r و c تستنفذ كل ما نهتم به من أشكال الصفة الأولى والصفة الثانية، على الترتيب ، وأنهما ليسا عينتين من عدد أكبر من الأشكال . أي أننا في هذا المثال أيضاً في حالة « نموذج الثوابت » أو « النموذج I » .

ونسوق الأمثلة التالية على حالة التصنيف الثنائي :

(أ) لدينا أطفال ثلاث مدارس (c = 3) وأربع طرق تعليم (r = 4) نريد إختبارها . ونفترض ، في « نموذج الثوابت » ، أن اهتمامنا لا يتعدى المدارس الثلاث ، وبصورة خاصة فإننا لا نستقرئ من النتائج أي شيء يتعلق بمجموعة أكبر من المدارس . ونريد هنا مقارنة المدارس ببعضها ومقارنة طرق التعليم ، وربما رؤية ما إذا كان يوجد أي تفاعل بين مدرسة وطريقة معينتين .

(ب) لدينا ثلاث مدن (c = 3) وقسمنا العاملين في كل منها إلى مجموعتين وفقاً للجنس (r = 2) . ونريد مقارنة متوسط دخل الرجال مع متوسط دخل النساء ، ومقارنة متوسط الدخل بين المدن الثلاث ، ثم رؤية ما إذا كان يوجد أي تفاعل كأن نجد ، مثلاً ، في مدينة معينة متوسط دخل كبير (أو صغير) بصورة غير اعتيادية للرجال أو للنساء ولكن ليس للجنسين معاً

(حـ) لدينا أربعة أنواع من القمح (c = 4) وثلاثة أنواع من الأسمدة (r = 3) ونريد اكتشاف ما إذا كان تركيب معين بين نوع من القمح ونوع من السماد يؤدي إلى متوسط إنتاج أفضل من التراكيب الباقية . أو أننا نريد معرفة ما إذا كان أحد أنواع الأسمدة أفضل ، على وجه الإجمال ، من النوعين الآخرين . (وتشير عبارة « على وجه الإجمال » إلى أننا نعتبر إنتاج هذا السماد مستخدماً مع الأنواع الأربعة من القمح) .

مثال ٣ ـ التصنيف المتعدد : ويمكن تعميم المثال السابق إلى حالات نصنف فيها الأفراد وفقاً لأكثر من متحولين . فمثلاً يمكن تصنيف سكان ثلاث مدن وفقاً لِـ : (أ) المدينة التي يعيشون فيها . (ب) الجنس ، و (حـ) كون الدخل أكثر أو أقل من 3600 ل.س سنوياً . وهكذا ينقسم السكان إلى 12 مجتمعاً جزئياً ، الرجال في المدينة 1 الذين يكسبون أكثر من 3600 ل.س سنوياً ، الرجال في المدينة 1 الذين يكسبون أقل من 3600 ل.س ؛ النساء في المدينة 1 اللواتي يكسبن أكثر من 3600 ل.س سنوياً الخ . ونريد معرفة ما إذا كان متوسط النسبة المئوية من الدخل السنوي ، الذي يُنفق على الطبابة ، نفسه : بالنسبة للمدن الثلاث ، بالنسبة لفئتي الدخل ، وبالنسبة للجنسين . أو ما اذا كان هناك نوع من التفاعل بين هذه التصنيفات .

مثال ٤ ـ مركبات التشتت : اذا كانت المجتمعات المدروسة وليكن عددها k ، مثلاً ، هي عينة من عدد أكبر من المجتمعات ، التي نرغب القيام باستقراء حولها ، ولا تشكل لوحدها هدفاً للدراسة ، فإننا نشير إلى مثل هذه الحالة « بنموذج مركبات التشتت » . ومع أن التحليل والعمليات الحسابية تبقى كما هي في « نموذج الثوابت » ، إلا أن تفسير النتائج يختلف في الحالتين . وبالإضافة إلى مسألة مقارنة المتوسطات $\mu_1, \mu_2, \dots, \mu_k$ للمجتمعات التي تطرقت إليها التجربة ، لدينا هنا مسألة الاستقراء إلى ما وراء هذه المجتمعات ، إلى الصف من المجتمعات الذي تشكل المجتمعات الـ k المدروسة عينة منه .

وكأمثلة على هذه الحالة نسوق ما يلي :

(أ) لدينا مجموعة من 50 مدرسة ونختار عينات من الطلبة من خمس من هذه المدارس (k=5) . ونريد إختبار فرضية حول متوسطات مقادير تتعلق بإنتاجية هذه المدارس مستخدمين المعلومات المتوفرة لنا من خمس من هذه المدارس اخترناها لهذه الغاية .

(ب) لدينا مجموعة من 200 مدينة ونختار عينات من سكان كل من عشر مدن (k=10) . ونريد تقدير متوسط الدخل أو تشتت متوسط الدخل في المائتي مدينة .

(حـ) يمكن استخدام سماد من أجل 10 أنواع من القمح . ونختار من بينها ثلاثة أنواع (k=3) ونريد تقدير متوسط الإنتاج عند إستخدام هذا السماد من أجل الأنواع العشرة . ونلاحظ هنا أن التصنيف أحادي بإعتبار أنه يوجد سماد واحد فقط .

مثال ٥ ـ مركبات متحولين والنموذج المختلط : تحصل حالة مركبات التشتت أيضاً في حالة تصنيف ثنائي أو متحولين . فإذا كانت الحالات التي يفترضها كل متحول أو صفة ضمن التجربة هي عينة من عدد أكبر من الحالات ، نقول إن النموذج المستخدم هو « نموذج مركبات التشتت » في حالة متحولين . وإذا استخدمنا في التجربة كل الحالات التي يفترضها أحد المتحولين وعينة فقط من الحالات التي يمكن أن يفترضها المتحول الآخر فنشير إلى مثل هذه الحالة على أنها حالة « النموذج المختلط » . وكأمثلة على هاتين الحالتين نذكر :

(أ) مركبات التشتت : لدينا مجموعة من 50 مدرسة وطلاب من 12 زمرة من زمر السنّ . نختار خمس مدارس (k=5) ، وزمرتين من زمر السنّ (r = 2) ، ونأخذ قياسات تتعلق بطلاب ضمن هاتين الزمرتين في كل

المدارس الخمس . ويمكن أن تتطرق القياسات إلى : القدرة على التعلّم ، والنقود المصروفة على وسائل التسلية ، معدل النموّ ، الخ . ونريد القيام باستقراء يتناول كل المدارس الخمسين وكل الزمر الاثنتي عشرة . وعلى سبيل المثال ، قد نرغب في معرفة ما إذا كان متوسط المهارة التعليمية في المدارس الخمسين فوق القياس المعتاد ، معتمدين في مثل هذه المعرفة على البيان الإحصائي الذي حصلنا عليه من المدارس الخمس .

(ب) مركبات التشتت : لدينا بمجموعة من المدن تحوي مائتي مدينة وخمسة عشر شريحة دخل بالنسبة للعاملين في هذه المدن . ونختار عينة من العاملين من كل من عشرين مدينة ($c = 20$) ، وذلك ضمن ثلاث شرائح للدخل ($r = 3$) نختارها ، ثم نقيس من أجل كل عامل تحصيله التربوي . ونريد القيام باستقراء يتعلق بمقارنة متوسطات التحصيل التربوي للعاملين في المائتي مدينة ، ومتوسطات التحصيل التربوي ضمن كل من شرائح الدخل الإثنتي عشرة . ويمكن أن نهتم أيضاً فيما إذا كانت هناك دلالة على وجود تفاعل بين المدينة ومستوى الدخل . فمثلاً يمكن أن توجد في بعض المدن شرائح دخل معينة يكون متوسط التحصيل التربوي فيها عالياً أو منخفضاً بصورة غير إعتيادية .

(ح) النموذج المختلط : لدينا أربعة أنواع من القمح ($c = 4$) وستين موقعاً عاماً في القطر العربي السوري نزرع فيه القمح . ونختار خمسة من هذه المواقع بصورة عشوائية ($r = 5$) ونزرع في كل منها الأنواع الأربعة من القمح . وبما أن كل أنواع القمح التي نهتم بدراستها موجودة في التجربة التي تحوي عينة من المواقع فقط (خمسة مواقع) فلدينا هنا تجربة نموذج مختلط .

وتحاول معظم التحريات العلمية أن تأخذ بعين الإعتبار عوامل أخرى إلى جانب العامل الذي هو موضع الدراسة . وغالباً ما تكون الطرق الموضحة

أعلاه ، والتي ترتب العوامل وفق تصنيفات عامة ، مفيدة لهذه الغاية ، بإعتبار أنها تسمح بالقيام بدراسة عند كل مستوى من مستويات كل من العوامل التي تضمها التجربة . ويأتي هذا النوع من التجارب العلمية في مقابل الطريقة التي تبقى فيها كل العوامل مثبتة بإستثناء العامل المدروس . ولإيضاح هذه النقطة نذكر الأمثلة التالية :

لنفرض أننا نرغب في مقارنة طريقتين في تعليم موضوع معين ونختار مجموعات من الطلبة فيها نفس نسبة الذكور والإناث ولأفرادها نفس المستوى من الذكاء ، ونفس الأعمار ، الخ . ولكن تنفيذ التجربة ، بمعلم واحد أو ضمن مدرسة واحدة ، يخضع للانتقاد بأن هذا المعلم بعينه يمتاز بكفاءة خاصة بالنسبة لإحدى الطريقتين ، أو أن تفوق إحدى الطريقتين يعود إلى المدرسة التي تمت فيها التجربة . ومن الواضح أنه ينبغي لتجربة من هذا النوع أن تحوي عدداً من المدارس وعدداً من المعلمين في كل مدرسة . وبالإستناد إلى طبيعة المسألة المدروسة ، فقد يكون من الأفضل أن تشمل التجربة مدارس أكثر ، حتى ولو اضطررنا إلى استخدام معلم واحد في كل مدرسة . وهكذا فإننا لا نرغب عند دراسة ظاهرة معينة أن نثبت كل العوامل بإستثناء العامل المدروس ، وإنما نرغب في تبيان وجود الظاهرة بصورة مستقلة عن العوامل الأخرى .

لنفرض الآن أننا نرغب في دراسة تأثير بضع معالجات على مجموعة من الفئران البيضاء . ولنفرض ، مثلاً ، أننا أعطينا زرقاً بتركيز مختلف إلى أربع مجموعات من الفئران ثم لاحظنا قدرتها على اجتياز متاهة . فإذا وجدنا أن هناك فرقاً يُذكر بين قدرات المجموعات الأربع ، ولكن اكتشفنا عند دراسة سجلاتها أنه توجد أيضاً فروق لا يمكن إغفالها بين هذه المجموعات تتعلق بأعمارها ، فلا يمكننا أن نعوّل على نتيجة مثل هذه التجربة ، بل نعتبرها باطلة . ومن السهل أن نرى بأن المطلوب هنا هو إزالة أية فرصة في أن يكون

لفروق في العمر أو التدريب أو الجنس أو فروق في تجارب سابقة خضعت لها هذه الفئران ، أثر أو قدرة على التغيير ، بحيث نستنتج خطأ أنها جاءت كنتيجة لمعالجاتنا ، أو أنها تسبب كثيراً من التغييرات في النتائج بحيث تحجب أية تأثيرات لمعالجاتنا .

واذا رغبنا في مقارنة إنتاجية بضع سلالات من الذرة الصفراء ، فلا بد من إعتبار عوامل هامة مثل وقت الزرع ، خصوبة التربة ، كمية السماد ، الخ . فإما أن يتمكن المجرب من التحكم بهذه المتحولات أو أن نصمم التجربة بحيث يمكن تقدير تأثيرات وقت الزرع ، خصوبة التربة ، الخ . ، وفصلها عن تأثير الفرق بين السلالات نفسها على الإنتاج . وقد يكون ضرورياً أن نستخدم عدة أوقات للزرع ، وأن نعالج الذرة الصفراء بكميات مختلفة من السماد ، الخ . لكي نستطيع تقدير مثل هذه التأثيرات .

ويجب أن تشمل دراسة طرق مختلفة لمعالجة أطفال قاصرين ، حالات من بيئات مختلفة ثقافياً واقتصادياً وجغرافياً الخ .

ويمكن اللجؤ إلى طرق إحصائية أبسط في الحالات التي يمكن فيها التحكم في المتحولات الهامة وإبقائها ثابتة ، كما هي الحال مثلاً في تجربة تجري داخل مخبر ، وعندما لا تسمح الظروف بالتحكم بمثل هذه المتحولات يجب تطوير طريقة إحصائية تأخذ بعين الاعتبار مثل هذه التغيرات المختلفة . وعلى أي حال فإنه حتى في الحالات التي يمكن التحكم فيها بكافة المتحولات الهامة ، قد يكون ذلك مكلفاً إلى الحد الذي يمنعنا عملياً من القيام به . بالإضافة إلى أنه قد يؤدي تثبيت هذه المتحولات الهامة عند قيم معينة إلى نتائج تجريبية مختلفة عما لو كنا ثبتناها عند قيم أخرى . وأحياناً تكون إعادة التجربة تحت نفس الشروط التجريبية المحيطة أمراً بالغ الصعوبة . . .

وعامل الزمن هو أحد العوامل الرئيسية التي تدعونا إلى تصميم تجربة تدرس عدداً من المعالجات أو العوامل في نفس الوقت . فغالباً ما تستغرق تجربة

زراعية أو بيولوجية عاماً كاملاً ، مما يجعل من الهام أن ننجز التجربة على أكمل وجه ممكن .

١١ ـ ٣ التصنيف الأحادي ـ النموذج I : في هذه الحالة ينتمي كل فرد إلى واحد وواحد فقط من k من المجتمعات المتميزة ، بمتوسطات هي $\mu_1 , \mu_2 , \ldots , \mu_k$. ونريد القيام باستقراءات تتعلق بمقادير هذه المتوسطات . وفي هذه الفقرة سنعرض طريقة لاختبار الفرضية بأن جميع هذه المتوسطات وعددها k متساوية فيما بينها .

نأخذ عينة عشوائية من كل مجتمع وليكن حجم كل منها n . ولنرمز لمجموع قياسات العينة الأولى بِـ T_1 ولمجموع قياسات العينة الثانية بِـ T_2 وهكذا ، ثم لنرمز للمجموع الكلي لجميع القياسات المأخوذة في التجربة بـ T . وبصورة مشابهة نرمز لمتوسط العينة الأولى بِـ $\bar{x}_1$ ومتوسط العينة الثانية بِـ $\bar{x}_2$ وهكذا ، ولمتوسط كل قياسات التجربة أو المتوسط الإجمالي بِـ $\bar{x}$. وبما أن هذه العينات تمثل ، في معظم التطبيقات العملية التي تهمنا ، نتائج تطبيق المعالجات المختلفة التي تدرسها التجربة فسنستخدم مصطلح « المعالجة » ليعني نفس ما تعنيه العينة . ويلخص الجدول (١١ ـ ٢) هذه الرموز .

جدول ١١ ـ ٢ حالة k من العينات حجم كل منها n .

	العينة الأولى أو المعالجة الأولى	**العينة الثانية أو المعالجة الثانية**	**...**	**العينة k أو المعالجة k**	
	x_{11}	x_{12}	...	x_{1k}	
	x_{21}	x_{22}	...	x_{2k}	
	$\vdots$	$\vdots$		$\vdots$	
	x_{n1}	x_{n2}	...	x_{nk}	
المجموع	T_1	T_2	...	T_k	المجموع الكلي T
المتوسط	$\bar{x}_1$	$\bar{x}_2$	...	$\bar{x}_k$	المتوسط الإجمالي $\bar{x}$

وترمز x_{ij} في متن الجدول إلى القياس i من العينة j أو نتيجة التطبيق i للمعالجة j .

ونعرف مجموع المربعات الكلي على الشكل :

$$SS = \sum_{i=1}^{n} \sum_{j=1}^{k} (x_{ij} - \bar{x})^2 \qquad (1)$$

ويعني المجموع المضاعف في هذه العلاقة أن الدليل الثاني j يتحول من 1 إلى k ويتحول الدليل الأول i فوق الأعداد الصحيحة من 1 إلى n . وبعبارة أخرى فإن مجموع المربعات الكلي هو مجموع مربعات انحرافات جميع القياسات أو الملاحظات التي تتضمنها التجربة عن متوسطها الإجمالي ؛ أي أنها تقيس مدى إنتشار أو تبعثر هذه الملاحظات حول متوسطها الإجمالي .

ونعرف مجموع مربعات ما بين المتوسطات على الشكل :

$$SST = n \sum_{j=1}^{k} (\bar{x}_j - \bar{x})^2 \qquad (2)$$

وباعتبار أن المتوسط الإجمالي $\bar{x}$ هو أيضاً متوسط المتوسطات $\bar{x}_1, \ldots, \bar{x}_k$ فإن العلاقة (2) تمثل مجموع مربعات إنحرافات متوسطات العينات عن المتوسط الإجمالي . وستنخفض قيمة هذا المجموع كلما كانت قيم متوسطات العينات قريبة من بعضها البعض . وعلى الوجه الآخر إذا تمخض تطبيق المعالجات عن تأثيرات مختلفة جداً فستكون هناك فروق كبيرة بين متوسطات العينات مما يؤدي إلى قيمة كبيرة لمجموع المربعات SST.

ونعرف مجموع مربعات ما ضمن العينات على الشكل :

$$SSE = \sum_{i=1}^{n} (x_{i1} - \bar{x}_1)^2 + \sum_{i=1}^{n} (x_{i2} - \bar{x}_2)^2 + \cdots + \sum_{i=1}^{n} (x_{ik} - \bar{x}_k)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{k} (x_{ij} - \bar{x}_j)^2 \qquad (3)$$

وفي المجموع الأول من الطرف الأيمن من العلاقة (3) نجد مجموع مربعات انحرافات قياسات العينة الأولى عن متوسطها وهو يقيس التشتت

ضمن العينة الأولى . وبصورة مشابهة يمثل الحد الثاني التشتت ضمن العينة الثانية وهكذا . ومجموع المربعات SSE هو قياس لتشتت الملاحظات ضمن العينات حول متوسطاتها الموافقة ، وهو مستقل عن أي فرق بين متوسطي عينتين ؛ وتعتمد قيمته فقط على التغيرات التصادفية التي ترافق أية تجربة وهي تمثل بالتالي الخطأ التجريبي .

مثال ٦ : لنعتبر العينات الخمس التالية وفي كل منها ثلاث ملاحظات :

	3	5	7	6	4	
	2	8	8	8	9	
	4	8	6	7	5	
المجموع	9	21	21	21	18	$T = 90$
المتوسط	3	7	7	7	6	$\bar{x} = 6$

وسنحسب في هذا المثال SS ، SST ، و SSE كما عرفناها أعلاه

$$SS = (3-6)^2 + (2-6)^2 + (4-6)^2 + (5-6)^2 + (8-6)^2 + \ldots + (5-6)^2$$
$$= 62$$

$$SST = 3\,[\,(3-6)^2 + (7-6)^2 + (7-6)^2 + (7-6)^2 + (6-6)^2\,]$$
$$= 3(9 + 1 + 1 + 1 + 0) = 36;$$

$$SSE = (3-3)^2 + (2-3)^2 + (4-3)^2 + (5-7)^2 + (8-7)^2 + (8-7)^2$$
$$+ \ldots + (5-6)^2 = 26$$

ولا نتبع عملياً هذه الطريقة في حساب SS ، SST و SSE إذ توجد طريقة أسهل بكثير للحصول على هذه المقادير ويمكن الاستفادة فيها من الآلات الحاسبة العادية ، وتعبر العلاقات التالية التي تسمى بالأشكال الحسابية لمجاميع المربعات المعرفة أعلاه ، عن هذه الطريقة :

$$SS = \sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij}^2 - \frac{T^2}{kn} \qquad (4)$$

$$SST = \frac{\sum_{j=1}^{k} T_j^2}{n} - \frac{T^2}{kn} \qquad (5)$$

$$SSE = \sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij}^2 - \frac{\sum_{j=1}^{k} T_j^2}{n} \qquad (6)$$

مثال ٢ : احسب SS ، SST ، و SSE في المثال 1 باستخدام الأشكال الحسابية .

$$SS = 3^2 + 2^2 + 4^2 + 5^2 + 8^2 + \dots + 5^2 - \frac{8100}{15} = 602 - 540 = 62;$$

$$SST = \frac{9^2 + 21^2 + 21^2 + 21^2 + 18^2}{3} - \frac{8100}{15} = 576 - 540 = 36;$$

$$SSE = 3^2 + 2^2 + 4^2 + 5^2 + 8^2 + \dots + 5^2 - \frac{9^2 + 21^2 + 21^2 + 21^2 + 18^2}{3}$$

$$= 602 - 576 = 26.$$

ولبيان صحة العلاقات (4) ، (5) و (6) يكفي أن نتذكر أنه اذا كان لدينا مجموعة من N من الأعداد $x_1, x_2, \dots, x_N$ متوسطها $\bar{x}$ فإن :

$$\sum_{i=1}^{N} (x_i - \bar{x})^2 = \sum_{i=1}^{N} x_i^2 - N\bar{x}^2 = \sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}$$

أي أن مجموع مربعات انحرافات N من الأعداد عن متوسطها يساوي إلى مجموع مربعات هذه الأعداد مطروحاً منها حاصل قسمة مربع مجموع هذه الأعداد على عددها . ومن العلاقة (1) نجد أن SS هو عبارة عن مجموع مربعات انحرافات القياسات الـ kn التي تحويها التجربة وهي $x_{11}, x_{21}, \dots, x_{kn}$ عن متوسطها ، وبالتالي فهي تساوي وفقاً للقاعدة السابقة مجموع مربعات هذه الأعداد $\sum_{i=1}^{N}\sum_{j=1}^{k} x_{ij}^2$ مطروحاً منها حاصل قسمة مربع مجموعها على عددها أي $\frac{T^2}{nk}$. وهكذا تكون العلاقة (4) هي شكل آخر للعلاقة (1) أسهل وأقصر بالنسبة للعمليات الحسابية .

وبتطبيق القاعدة السابقة على الأعداد $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ ومتوسطها $\bar{x}$ نجد أن العلاقة (2) تصبح :

$$SST = n\left[\sum_{j=1}^{k} \bar{x}_j^2 - \frac{\left(\sum_1 \bar{x}_j\right)^2}{k}\right]$$

$$= n\left[\sum_{j=1}^{k} \frac{T_j^2}{n^2} - \frac{\left(\sum_{j=1}^{k} T_j\right)^2}{k n^2}\right] = \frac{\sum_{j=1}^{k} T_j^2}{n} - \frac{T^2}{kn}$$

وهي العلاقة (5) .

ولبيان أن (6) هي شكل آخر للعلاقة (3) نطبق القاعدة المذكورة أعلاه على كل من مجاميع المربعات الـ k التي تظهر في (3) فنجد :

$$SSE = \sum_{i=1}^{n} x_{i1}^2 - \frac{\left(\sum_{i=1}^{n} x_{i1}\right)^2}{n} + \sum_{i=1}^{n} x_{i2}^2 - \frac{\left(\sum_{i=1}^{n} x_{i2}\right)^2}{n}$$

$$+ \cdots + \sum_{i=1}^{n} x_{ik}^2 - \frac{\left(\sum_{i=1}^{n} x_{ik}\right)^2}{n}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij}^2 - \frac{T_1^2}{n} - \frac{T_2^2}{n} - \cdots - \frac{T_k^2}{n} = \sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij}^2 - \frac{\sum_{j=1}^{k} T_j^2}{n}$$

ويلاحظ القارئ من المثال (1) أن مجموع SST و SSE هو SS أي أن 62 = 26 + 36، وفي الحقيقة فإن هذه الخاصة هي قاعدة تصح من أجل أية مجموعة من القياسات ولبيان ذلك نكتب :

$$SST + SSE = \frac{\sum_{j=1}^{k} T_j^2}{n} - \frac{T^2}{kn} + \sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij}^2 - \frac{\sum_{j=1}^{k} T_j^2}{n}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij}^2 - \frac{T^2}{kn} = SS$$

واذا كان $\overline{x} = 0$ فإن SS تصبح مساوية لمجموع مربعات القياسات في التجربة أي $\sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij}^2$، وهو الحد الأول من العلاقة (4). ولذلك فإن الحد $\frac{T^2}{kn}$ يلعب دور المصحح بحيث يصبح مجموع المربعات مأخوذاً حول المتوسط وليس حول الصفر، وهذا هو السبب في أن الحد $\frac{T^2}{kn}$ يدعى

في كتب الإحصاء التطبيقي حدّ التصحيح وسنرمز له بـ C أي أن :

$$C = \frac{T^2}{kn}$$

وسنقول بعد الآن مجموع المربعات حول المتوسط لتعني مجموع مربعات الإنحرافات عن المتوسط .

وإذا لم يكن للعينات الـ k نفس الحجم ، أي لو فرضنا أن حجم العينة الأولى n_1 وحجم العينة الثانية n_2 وهكذا ، فإن الجدول (١١ ـ ٢) يصبح على الشكل :

جدول ١١ ـ ٣ حالة k عينة بأحجام غير متساوية .

	العينة الأولى أو المعالجة الأولى	العينة الثانية أو المعالجة الثانية		العينة k أو المعالجة k	
	x_{11}	x_{12}	...	x_{1k}	
	x_{21}	x_{22}	...	x_{2k}	
	$x_{n_1,1}$	$x_{n_2,1}$	...	$x_{nk,k}$	
المجموع	T_1	T_2	...	T_k	المجموع الكلي T
المتوسط	$\bar{x}_1$	$\bar{x}_2$	...	$\bar{x}_k$	المتوسط الإجمالي $\bar{x}$

وتحتاج العبارات المعرفة بالعلاقات (1) ، (2) و (3) إلى تعديلات بسيطة لتصبح متفقة مع الحالة الجديدة . وهنا تأخذ العلاقات الأشكال التالية :

$$SS = \sum_{i=1}^{n_1}(x_{i1}-\bar{x})^2 + \sum_{i=1}^{n_2}(x_{i2}-\bar{x})^2 + \cdots + \sum_{i=1}^{n_k}(x_{ik}-\bar{x})^2$$
$$= \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij}-\bar{x})^2 \quad (7)$$

$$SST = n_1(\bar{x}_1-\bar{x})^2 + n_2(\bar{x}_2-\bar{x})^2 + \cdots + n_k(\bar{x}_k-\bar{x})^2$$
$$= \sum_{j=1}^{k} n_j(\bar{x}_j-\bar{x})^2 \quad (8)$$

$$SSE =$$

$$\sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1)^2 + \sum_{i=1}^{n_2} (x_{i2} - \bar{x}_2)^2 + \cdots$$

$$+ \sum_{i=1}^{n_k} (x_{ik} - \bar{x}_k)^2 = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_k)^2 \quad (9)$$

وتوجد بصورة مشابهة أشكال حسابية لهذه العلاقات :

$$SS = \sum_{i=1}^{n_1} x_{i1}^2 + \sum_{i=1}^{n_2} x_{i2}^2 + \cdots \sum_{i=1}^{n_k} x_{ik} - \frac{T^2}{n_1 + n_2 + \cdots + n_k}$$

$$= \sum_{j=1}^{k} \sum_{i=1}^{n_j} x_{ij}^2 - \frac{T^2}{\sum_{j=1}^{k} n_j} \qquad (10)$$

$$SST = \frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \cdots + \frac{T_k^2}{n_k} - \frac{T^2}{n_1 + n_2 + \cdots + n_k}$$

$$= \sum_{j=1}^{k} \frac{T_j^2}{n_j} - \frac{T^2}{\sum_{j=1}^{k} n_j} \qquad (11)$$

$$SSE = \sum_{i=1}^{n_1} x_{i1}^2 + \sum_{i=1}^{n_2} x_{i2}^2 + \cdots + \sum_{i=1}^{n_k} x_{ik}^2 - \left(\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \cdots + \frac{T_k^2}{n_k}\right)$$

$$= \sum_{j=1}^{k} \sum_{i=1}^{n_j} x_{ij}^2 - \sum_{j=1}^{k} \frac{T_j^2}{n_j} \qquad (12)$$

لنعد الآن إلى حالة عينات متساوية الحجم ولنفرض أننا نريد إختبار الفرضية بأن هذه العينات جاءت من مجتمع واحد ، وبعبارة أوضح نريد إختبار الفرضية $\mu_1 = \mu_2 = \cdots = \mu_k$ فعندئذ يمكن تقدير التشتت في هذا المجتمع ولنرمز له بـ σ^2 بطريقتين والحصول على تقديرين مستقلين تماماً لِـ σ^2 .

١ ـ يمكن تقدير σ^2 بضم المعلومات من العينات الـ k ، وذلك كتعميم للطريقة التي استخدمناها في حالة مقارنة متوسطين ، والتي درسناها في فصلين سابقين . وهكذا نجد في الصورة مجموع المربعات حول المتوسط ضمن كل من العينات الـ k ، وفي المخرج نجد $n_1 + n_2 + \dots + n_k - k$ ، وفي حالة تساوي حجوم العينات أي $n_1 = n_2 = \dots = n_k = n$ نجد في المخرج $nk-k=(n-1)k$ وهكذا يكون التقدير الذي نحصل عليه بهذه الطريقة هو :

$$S_p^2 = \frac{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2 + \sum_{i=2}^{n}(x_{i2}-\bar{x}_2) + \dots + \sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2}{k(n-1)}$$

$$= \frac{\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij}-\bar{x}_j)^2}{k(n-1)} \qquad (13)$$

ولكن الكمية في الصورة هي SEE وهكذا نكتب :

$$S_p^2 = \frac{SSE}{k(n-1)} \qquad (14)$$

٢ ـ ويمكن الحصول على التقدير الآخر لِـ σ^2 بأن ننظر إلى العينات الـ k على أنها مسحوبة ، وفقاً للفرضية التي انطلقنا منها ، من نفس المجتمع . ومتوسطات هذه العينات $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ تمثل k من القياسات التي افترضها متحول $\bar{x}$ تشتته هو $\sigma_{\bar{x}}^2 = \sigma^2/n$ ، تماماً كما رأينا في الجدول (٦ ـ ١) في الجزء الأول الذي يحوي عينات من مجتمع قذف قطعة زهر . وتمثل القيم $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ إذن عينة حجمها k من المجتمع الذي يمثل كل القيم الممكنة لِـ $\bar{x}$ (متوسط عينة حجمها n نسحبها من المجتمع المدروس) . وتشتت هذه العينة من المتوسطات هو :

$$S_{\bar{x}}^2 = \frac{\sum_{j=1}^{k}(\bar{x}_j-\bar{x})^2}{k-1}$$

وهو تقدير لِـ $\frac{\sigma^2}{n}$ وهكذا نجد التقدير الآخر لِـ σ^2 على أنه $n S_{\bar{x}}^2$ وسنرمز له بِـ S_t^2 أي أن :

$$S_t^2 = \frac{n \sum_{j=1}^{k} (\bar{x}_j - \bar{x})^2}{k-1} \qquad (15)$$

ولكن الصورة في هذه العلاقة هي SST وهكذا نكتب التقدير الثاني المستقل لِ σ^2 على الشكل :

$$S_t^2 = \frac{SST}{k-1} \qquad (16)$$

وتتألف طريقة تحليل التشتت من مقارنة هذين التقديرين المستقلين لِ σ^2. وباعتبارهما تقديران لنفس الكمية فإن نسبتهما يجب ألا تختلف كثيراً عن الواحد . وفي الحقيقة يُبرهن في الاحصاء الرياضي على ما يلي :

إذا كررنا مثل هذه التجربة ، أي أخذنا k عينة حجم كل منها n من مجتمع طبيعي ، عدداً كبيراً من المرات ، وحسبنا في كل مرة النسبة S_t^2 / S_p^2 حيث S_t^2 و S_p^2 معرفان بالعلاقتين (16) و (14) على الترتيب ، فإن التوزيع الإحتمالي لهذه النسبة يتبع التوزيع F (المعرف في الفقرة ٦ ــ ٦) بِـ $\nu_1 = k - 1$ و $\nu_2 = k(n-1)$ درجة من الحرية .

والجدير بالملاحظة هو أن مجموع درجات الحرية الموافقة لِ S_t^2 وهو $(k-1)$ ، وعدد درجات الحرية الموافقة لِ S_p^2 وهو $k(n-1)$ ، يساوي $(k-1) + k(n-1) = kn-1$ ، وهو عدد درجات الحرية الموافقة للعدد الكلي للقياسات التي تحويها التجربة . وهكذا نجد أن تحليل أو تفكيك التشتت يرافقه تفكيك موافق لعدد درجات الحرية .

والآن إذا لم تكن فرضيتنا صحيحة ، أي اذا كانت العينات الـ k مأخوذة في الحقيقة من مجتمعات تختلف في متوسطاتها ، فإن S_t^2 ستختلف كثيراً عن S_p^2 نظراً للتشتت الأوسع لمتوسطات العينات حول المتوسط الإجمالي . وإذا كان S_t^2 كبيراً بالمقارنة مع S_p^2 فإن قيمة النسبة F الناتجة ستتجاوز $F_{1-\alpha}$، وعندها سنرفض الفرضية ، ونستنتج أنه يوجد على الأقل زوج من

المتوسطات $\mu_1, \mu_2, \cdots, \mu_k$ تختلف عن بعضها إختلافا لا يمكن إهماله . ولنلاحظ أننا نصل إلى إستنتاج وجود فروق هامة بين المتوسطات ، فقط في الحالة التي يكون فيها S_t^2 أكبر من S_p^2 ، أي اذا كانت النسبة $F=S_t^2/S_p^2$ تتجاوز الواحد ، وهكذا فإننا نستخدم على الدوام إختباراً وحيد الذيل في تحليل التشتت . وعندما يكون S_t^2 أقل من S_p^2 نصل سلفاً إلى قبول الفرضية ، وليس من الضروري أن نحسب النسبة F أو نعقد مقارنة مع F_α .

ولاختبار الفرضية $\mu_1=\mu_2=\cdots=\mu_k$ بطريقة تحليل التشتت نقوم بالخطوات التالية :

١ ــ نحسب حد التصحيح : $C = \dfrac{T^2}{kn}$

٢ ــ نحسب مجموع المربعات الكلي $SS = \sum_{i=1}^{n}\sum_{j=1}^{k} X_{ij}^2 - C$

٣ ــ نحسب مجموع مربعات المعالجات (ما بين المتوسطات)

$$SST = \frac{\sum_{i=1}^{k} T_i^2}{n} - C$$

٤ ــ نحسب مجموع مربعات الخطأ $SSE = SS - SST$

٥ ــ نحسب عدد درجات الحرية الموافق لكل من مجاميع المربعات السابقة .

٦ ــ نحسب $S_t^2 = \dfrac{SST}{k-1}$. ونرمز لهذه الكمية عادة بِـ MST .

٧ ــ نحسب $S_p^2 = \dfrac{SSE}{k(n-1)}$. ونرمز عادة لهذه الكمية بِـ MSE.

٨ ــ نحسب النسبة $F = MST/MSE$

٩ ــ نقارن القيمة الناتجة لِـ F مع $F_{1-\alpha}((k-1, k(n-1))$ كما نجدها من جدول التوزيع F ونرفض الفرضية إذا كانت F أكبر من $F_{1-\alpha}$ وذلك عند مستوى الأهمية α .

ونلخص هذه الخطوات عادة في جدول يسمى جدول تحليل التشتت وذلك على الشكل التالي :

جدول ١١ ـ ٣ تحليل التشتت في حالة عينات متساوية الحجم

مصدر التغير	مجموع المربعات	درجات الحرية	متوسط المربعات	F
ما بين العينات (أو المعالجات)	$SST = \frac{\sum^{k} T_{j}^{2}}{n} - C$	$k - 1$	$MST = \frac{SST}{k-1}$	
ما ضمن العينات (أو الخطأ)	$SSE = SS - SST$	$k(n-1)$	$MSE = \frac{SSE}{k(n-1)}$	$\frac{MST}{MSE}$
المجموع	$SS = \sum\sum x_{ij}^{2} - C$	$kn - 1$		

١١ ـ ٤ **التصميم التام العشوائية** : يبرز التصميم التام العشوائية من الطريقة التي نخطط فيها تجربة من تجارب التصنيف الأحادي . فلنفرض أن لدينا خمسة أنواع من الأسمدة ونرغب في إختبار الفرضية الإبتدائية بعدم وجود فرق بين تأثيرات هذه الأسمدة على إنتاج الذرة الصفراء . وسنفرض أنه تتوفر لنا 20 وحدة تجريبية متجانسة بالنسبة لمختلف الشروط التي تؤثر في الإنتاج . والطريقة السليمة لتخصيص كل سماد لأربع من هذه الوحدات التجريبية (قطع أرض صغيرة) هو أن تتم عملية التخصيص هذه بصورة عشوائية . ويمكن إتمام ذلك بترقيم الوحدات التجريبية من واحد إلى عشرين . ثم نضع عشرين قطعة صغيرة من الورق في قبعة ، مثلاً ، وعلى كل أربع منها علامة توافق أحد الأسمدة الخمسة (رقم أو لون أو كتابة اسم السماد) . ونخلط هذه الأوراق جيداً ثم نسحب بصورة عشوائية واحدة منها فتحدد لنا نوع السماد الذي سنطبقه في الوحدة التجريبية الأولى ونسحب الثانية فتحدد السماد الذي سنطبقه في الوحدة التجريبية الثانية ، وهكذا . وفي مثل هذه الحالة نقول أننا نصمم التجربة وفق التصميم التام العشوائية ، بإعتبار أنه يمكن تخصيص أي معالجة لأي من الوحدات التجريبية المتوفرة . وبعد إنجاز التجربة عملياً ، نحصل أخيراً على إنتاج الذرة الصفراء من كل من الوحدات التجريبية العشرين . ولو أننا خصصنا للأسمدة المختلفة أعداداً مختلفة من قطع الورق العشرين في القبعة ، لحصلنا على خطة للتجربة لا نطبق فيها كلاً من الأسمدة الخمسة على نفس العدد من الوحدات التجريبية ، وهذا يوافق ما سميناه أعلاه بحالة عينات غير متساوية الحجم .

مثال ٧ : لنفرض أن أرقام الإنتاج في التجربة المذكورة أعلاه ، وهي إختبار فعالية خمسة أنواع من الأسمدة في إنتاج الذرة الصفراء ، كانت كما يلي :

جدول ١١-٤ إنتاج عشرين وحدة تجريبية عولجت بخمس أنواع من الأسمدة

	1	2	3	4	5
	40	38	44	41	34
	45	40	42	43	35
	46	38	40	40	34
	49	44	34	40	33
المجموع الكلي 800	180	160	160	164	136

باتباع الخطوات المذكورة أعلاه في الفقرة (١١-٣) نجد بإعتبار $n = 4$ و $k = 5$ أن :

١ - $C = \frac{T^2}{nk} = \frac{(800)^2}{20} = 32000$

٢ - $SS = \sum_{i=1}^{4} \sum_{j=1}^{5} X_{ij}^2 - C = (40)^2 + (45)^2 + \ldots$

$+ (33)^2 - 32000 = 32378 - 32000 = 378$

٣ - $SST = \frac{1}{4} \sum_{j=1}^{5} T_j^2 - C = \frac{1}{4} [(180)^2 + (160)^2 + (160)^2$
$+ (164)^2 + (136)^2] - C = 32248 - 32000 = 248$

٤ - $SSE = SS - SST = 378 - 248 = 130$

ويكون جدول تحليل التشتت كما يلي :

جدول ١١-٥ تحليل التشتت للتجربة المعطاة في الجدول ١١ - ٤ .

مصدر التغير	درجات الحرية	مجموع المربعات	متوسط المربعات	F
ما بين الأسمدة	4	248	62	
الخطأ	15	130	8.67	$\frac{62}{8.67} = 7.15$
المجموع	19	378		

ولإختبار الفرضية بأن $\mu_1 = \mu_2 = \cdots = \mu_5$ عند المستوى $\alpha = .01$ نقارن F بـ $F_{.01}(4,15) = 4.89$. وبما أن $F > F_{.01}$ فإننا نرفض الفرضية عند المستوى $\alpha = .01$. ونستنتج وجود فروق بين تأثيرات الأسمدة الخمسة في الإنتاج .

ولظروف تتعلق بالتجربة وطبيعة المعالجات ، فقد تكون حجوم العينات الـ k غير متساوية . وفي هذه الحالة يصبح جدول تحليل التشتت كما هو مبين في الجدول (١١-٦) .

حيث حد التصحيح C هو $C = T^2 / \sum_{j=1}^{k} n_j$

مثال ٨ : لنفرض أننا نريد مقارنة تأثير أربعة أنواع من الأسمدة على إنتاج القمح وكان الإنتاج محسوباً بالبوشل / إيكر كما هو مبين في الجدول (١١-٧) .

جدول ١١-٦ تحليل التشتت في حالة عينات غير متساوية الحجوم .

مصدر التغير	درجات الحرية	مجموع المربعات	متوسط المربعات	F
ما بين المعالجات	$k-1$	$SST=\sum^{k}\frac{T_j^2}{n_j}-C$	$MST=\frac{SST}{k-1}$	
الخطأ	$\sum^{k}(n_j-1)$	SSE	$MSE=\frac{SSE}{\sum(n_j-1)}$	
المجموع	$\sum^{k}n_j-1$	$SS=\sum^{k}\sum^{k}x_{ij}^2-C$		

جدول ١١ـ٧ نتائج التجربة في المثال ٨ .

السماد

	1	2	3	4
	45	35	34	41
	46	33	34	41
	49		35	44
	44		34	43
			33	41
				42
				44
				41
				41
المجموع الكلي 800	184	68	170	378

وباتباع نفس الخطوات المذكورة أعلاه حيث $n_1 = 4$، $n_2 = 2$، $n_3 = 5$، $n_4 = 9$، $k = 4$، نجد :

١ ـ $$C = T^2 / \sum_{j=1}^{4} n_j = \frac{(800)^2}{20} = 32000$$

٢ ـ $$SS = \sum_{j=1}^{4} \sum_{i=1}^{n_j} x_{ij}^2 - C = (45)^2 + (46)^2 + \ldots + (41)^2 - C = 32464 - 32400 = 464$$

$$SST = \sum_{j=1}^{4} \frac{T_j^2}{n_j} - C = \frac{(184)^2}{4} + \frac{(68)^2}{2} + \frac{(170)^2}{5} + \frac{(378)^2}{9} - C = 32432 - 32000 = 432$$

٤ ـ $$SSE = SS - SST = 464 - 432 = 32$$

ويكون جدول تحليل التشتت هو :

جدول ١١ ــ ٨ تحليل التشتت للتجربة المعطاة في الجدول ١١ ــ ٧

مصدر التغير	درجات الحرية	مجموع المربعات	متوسط المربعات	F
ما بين الأسمدة	3	432	144	$\frac{144}{2} = 72$
الخطأ	16	32	2	
المجموع	19	464		

ولاختبار الفرضية بأن $\mu_1=\mu_2=\mu_3=\mu_4$ ، أي أنه لا فرق بين تأثيرات الأسمدة الأربعة على إنتاج القمح ، عند المستوى $\alpha=.01$، نقارن F كماحسبناها من الجدول(١١ـ٨) بـ $F_{.01}\ (3,16) = 5.29$ وبما أن $F > F_{.01}$ فإننا نرفض الفرضية عند المستوى $\alpha = .01$ (وهذا يعني أننا نرفضها حكماً عند المستوى $\alpha = .05$) .

١١ ــ ٥ **الفروض القائمة وراء طرق تحليل التشتت** : بما أن للفهم الواضح للشروط المختلفة التي تعتمد عليها طرق تحليل التشتت ، أهميته القصوى في مجال تفسير المعلومات التجريبية ، وتفسير نتائج التحليل وفي مجال التطبيق السليم لهذه الطرق ، فإننا سنلقي فيما يلي نظرة فاحصة على مثل هذه الشروط ، التي يجب أن ندرك بأنها ضرورية ، حتى يكون تحليلنا دقيقاً ، وحتى يكون من الممكن أن تترافق الاستقراءات التي نقوم بها بعبارات إحتمالية تحدد مدى الثقة بهذه الاستقراءات . وبالطبع فإن مثل هذه الفروض يجب أن تكون من طبيعة رياضية . وسنتطرق هنا للنموذجين I و II في حالة تصنيف أحادي ، ويمكن تعميمها بسهولة إلى حالات تصنيف ذي بعدين أو أكثر . أي أننا سنفترض هنا أننا نهتم بتقدير أو اختبار تأثيرات عامل واحد فقط ، وكل العوامل الأخرى يشملها الحد الذي نسميه حد « الخطأ » ،

والذي يتضمن كل مصادر التغير التي تقع خارج حدود العامل المدروس .

النموذج ا ـ نفترض ما يلي :

١ ـ الملاحظات x_{ij} هي القيم الملحوظة لمتحولات عشوائية تتوزع حول متوسط ξ_j $(j = 1, 2, ..., k)$

٢ ـ يمكن التعبير عن كل وسيط (وهو عدد ثابت) على الشكل :

$$\xi_j = \mu + \alpha_j \qquad (17)$$

وهي خاصة التجميعية حيث :

$$\mu = \frac{\sum_{j=1}^{k} \xi_j}{k} \qquad , \qquad \alpha_j = \xi_j - \mu$$

وهكذا يكون :

$$\sum_{j=1}^{k} \alpha_j = 0 \qquad (18)$$

٣ ـ للمتحولات العشوائية x_{ij} نفس التشتت σ^2 (خاصة التجانس) .

٤ ـ تتوزع المتحولات x_{ij} ، مستقلة عن بعضها ، وفق التوزيع الطبيعي .

ويمكن التعبير عن هذه الخواص باختصار كما يلي : يمكن التعبير عن الملاحظات وفقاً للنموذج الرياضي :

$$x_{ij} = \mu + \alpha_j + \varepsilon_{ij} \qquad \begin{matrix} i = 1, ..., n, \\ j = 1, ..., k. \end{matrix} \qquad (19)$$

حيث μ و α_j أعداد ثابتة وحيث : $\sum_{j=1}^{k} \alpha_j = 0$

والمتحولات ε_{ij} هي متحولات عشوائية مستقلة فيما بينها ويتوزع كل منها وفق التوزيع الطبيعي بمتوسط يساوي الصفر وبتشتت مشترك σ^2 ، وتحت هذه الشروط يصبح تطبيق طرق تحليل التشتت لتقدير وإختبار أهمية تأثيرات « المعالجات » أمراً مشروعاً بالإضافة إلى أنه يمكن وضع مقاييس لمدى الثقة بالاستقراءات التي نستخلصها من هذا التحليل .

النموذج II ـ مركبات التشتت ـ نفترض ما يلي :

١ ـ الملاحظات X_{ij} $(i = 1, ..., n; j = 1, ..., k)$ هي القيم الملحوظة لمتحولات عشوائية تتوزع حول متوسط يساوي μ .

٢ ـ كل متحول عشوائي x_{ij} هو مجموع مركبتين عشوائيتين (أي متحولين عشوائيين) والعلاقة التابعية بين x_{ij} وهذه المركبات هي :

$$x_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

حيث α_j و ε_{ij} هما متحولان عشوائيان .

٣ ـ يتوزع المتحولان العشوائيان α_j و ε_{ij} وفق التوزيع الطبيعي بمتوسط يساوي الصفر ، ولكن بتشتتين هما ، على الترتيب ، σ_α^2 و σ^2

٤ ـ المتحولات α_j ، $(j = 1, ..., k)$ و ε_{ij} $(i = 1, ..., n; j = 1, ..., k)$ هي متحولات مستقلة فيما بينها ، وتتوزع وفق التوزيع الطبيعي .

وبصورة عامة نقول أن شروط : التوزع الطبيعي ، التجميعية ، الاستقلال ، والتجانس تشكل الأساس الذي تقوم عليه الطرق الإحصائية العامة التي يستخدمها الباحثون العلميون .

وقد دُرست التأثيرات الناتجة عن عدم تحقق هذه الشروط تماماً من قبل كوكران وآخرون ، فوجدوا أن انحرافاً معتدلاً عن هذه الشروط سوف لا يثير صعوبات جدية ، ولكن اذا كانت هذه الإنحرافات كبيرة فيجب أن ننظر بعين الحذر للتحليل وللنتائج المستخلصة . وبعبارات أوضح نقول إن أكثر ما يسبب الصعوبات الجدية في طريق التحليل هو:أن يكون توزيع المتحولات بعيداً جداً عن التناظر ، تواجد الأخطاء بحجم كبير (سواء أكان الخطأ ناتجاً عن عدم دقة القياسات ، أو إغفال عوامل مؤثرة في النتائج لم يأخذها المجرب بعين الإعتبار ، أو إختيار التصميم غير المناسب للتجربة) ، ابتعاد النموذج الرياضي عن الشكل التجميعي ، والتغيرات في تشتت الخطأ ،

أي عدم التجانس ، (سواء أكان ذلك ناشئاً عن المتوسط أو عن معالجات بعينها ، أو من أجزاء من التجربة) . وأفضل الطرق لتحسين الوضع هو حذف بعض الملاحظات أو المعالجات أو تكرارات التجربة ، تجزئة خطأ التشتت ، أو التحويل إلى سلم قياس آخر قبل الشروع في التحليل . وعلى أي حال فإن إختيار الطريقة الأدق تتطلب خبرة كبيرة في طرق تحليل التشتت المتعددة . وسيكون من المستحسن غالباً أن يعود المجرب في مثل هذه الحالات إلى إختصاصي في الإحصاء الرياضي .

١١ ـ ٦ اختبار « بارتلت » من أجل تجانس التشتتات : سنتوقف هنا قليلاً لإعطاء إختبار هام نستطيع بواسطته الحكم على مدى إنسجام البيان الإحصائي المطلوب تحليله مع خاصة التجانس المذكورة في الفقرة السابقة . فقد توصل « بارتلت » إلى طريقة لإختبار الفرضية $H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$ حيث $\sigma_1^2, \ldots, \sigma_k^2$ هي تشتتات المجتمعات الـ k التي تهدف طريقة تحليل التشتت إلى مقارنة متوسطاتها . وذلك باستخدام k عينة ، واحدة من كل من هذه المجتمعات ، أحجامها بصورة عامة $n_1, n_2, \ldots, n_k$. ولنتذكر أن هذه المجتمعات هي بالفرض مجتمعات طبيعية . ونلخص الطريقة في الجدول (١١ـ ٩) الذي نستنتج منه قيمة χ^2 ثم نقارن هذه القيمة مع χ^2_α بـ $(k-1)$ درجة من الحرية كما نجده من جدول التوزيع χ . ونرفض الفرضية H_0 اذا كان $\chi^2 > \chi^2_\alpha$ ونستنتج في هذه الحالة أن المعلومات المطلوب تحليلها لا تحقق خاصة التجانس . وسيجد الباحث أنه من الضروري أن نحسب القيمة المصححة لِ χ ، فقط في الحالة التي تتجاوز فيها قيمة χ^2 غير المصحح الكمية χ^2_α وتكون بنفس الوقت قريبة منها ، أي أن مقدار التجاوز طفيف من جهة ، ويرغب المجرب في الحصول على تقييم دقيق لحجم الخطأ من النوع الأول α من جهة أخرى .

جدول ١١ـ ٩ حسابات اختبار « بارتلت » الخاص بتجانس التشتتات

العينة	مجموع مربعات العينة	درجات الحرية	مقلوب درجات الحرية	S_j^2 متوسط المربعات	$\log_{10} S_j^2$	(درجات الحرية) X ($\log_{10} S_j^2$)
1	$\sum_{i=1}^{n_1} X_{ij}^2 - \frac{T_1^2}{n_1}$	$n_1 - 1$	$1/(n_1-1)$	S_1^2	$\log_{10} S_1^2$	$(n_1-1) . \log_{10} S_1^2$
2	$\sum_{i=1}^{n_2} X_{ij}^2 - \frac{T_2^2}{n_2}$	$n_2 - 1$	$1/(n_2-1)$	S_2^2	$\log_{10} S_2^2$	$(n_2-1) . \log_{10} S_2^2$
⋮						
k	$\sum_{i=1}^{n_k} X_{ij}^2 - \frac{T_k^2}{n_k}$	$n_k - 1$	$1/(n_k-1)$	S_k^2	$\log_{10} S_k^2$	$(n_k-1) . \log_{10} S_k^2$
المجموع	W_{yy}	$\sum_{j=1}^{k} (n_j-1)$	$\sum_{j=1}^{k} \frac{1}{n_j-1}$	- - -	- - -	$\sum_{j=1}^{k} (n_j-1) \log_{10} S_j^2$

التقدير المركب للتشتت من جميع العينات = $S^2 = \dfrac{W_{yy}}{\sum_{j=1}^{k}(n_j-1)}$

$$B = (\log_{10} S^2)\sum_{j=1}^{k}(n_j-1)$$

$$\chi^2(k-1) = \log_e 10\left[B - \sum_{j=1}^{k}(n_j-1)\log_{10} S_j^2\right] \qquad (20)$$

عامل التصحيح $C = 1 + \dfrac{1}{3(k-1)}\left[\sum_{j=1}^{k}\dfrac{1}{n_j-1} - 1/\sum_{j=1}^{k}(n_j-1)\right]$

$\chi^2(k-1)$ مصحح $= \dfrac{1}{C}\chi^2(k-1)$ لاحظ أن: $\log_e 10 = 2.3026$

وللتأكد من قدرة القارئ على استخدام الرموز في الجدول (١١-٩) نأخذ المثال العددي المبين في الجدول (١١-١٠) . ونتائج الحسابات الضرورية مبينة في الجدول (١١-١١) ونرى أننا لا نستطيع رفض الفرضية بأن التشتتات متجانسة . ويلاحظ القارئ أنه لا ضرورة في هذا المثال لحساب χ^2 المصحح . ومع هذا فقد قمنا بالحسابات فقط ليكون المثال كاملاً .

جدول ١١ ـ ١٠ أربع عينات من مجتمعات طبيعية

A	B	C	D
48	42	33	78
49	39	42	69
67	51	46	60
75	57	47	52
53	75	50	63
33			45
			50
			35

جدول ١١ـ١١ حسابات إختبار « بارتلت » للبيان الإحصائي في الجدول ١١ـ١٠

العينة	مجموع مربعات العينة	درجات الحرية	مقلوب درجات الحرية	s_i^2	$bog_{10}\ s_i^2$	درجات الحرية x $(Log_{10}\ s_i^2)$
1	1113.0	5	.20	222.6	2.34753	11.73765
2	820.8	4	.25	205.2	2.31218	9.24872
3	173.2	4	.25	43.3	1.63649	6.54596
4	1330.0	7	.1428	190.0	2.27875	15.95125
المجموع	3437.0	20	.8428	. . . . .	. . . . .	43.48358

التقدير المركب للتشتت $= S^2 = \frac{3737}{20} = 171.85$

$$B = (\log_{10} S^2) \sum_{j=1}^{4} (n_j - 1) = (2.23515)\ (20) = 44.7030$$

$$\chi^2(3) = (2.3026)\ (44.7030 - 43.48358) = 2.80784$$

عامل التصحيح : $C = 1 + \frac{1}{3\,(3)} \left(.8428 - \frac{1}{20}\right) = 1.0881$

$\chi^2(3)$ المصحح $= \frac{2.80784}{1.0881} = 2.5805$

١١ - ٧ توقع متوسط المربعات في التصميم التام العشوائية : قبل اختبار أية فرضية تتعلق بتأثير « معالجات » عند تحليل بيان احصائي ، فإن شروطاً معينة يجب أن تتوفر وهي الشروط التي ذكرناها في الفقرة (١١ - ٥) . وفي حالتنا الخاصة في المثال ، إذا سمينا الأسمدة بالمعالجات فيمكن كتابة النموذج الرياضي على الشكل :

$$x_{ij} = \mu + \tau_j + \varepsilon_{ij}$$

$$i = 1, \ldots, n_j$$
$$j = 1, \ldots, 4$$

حيث τ_j هو التأثير الحقيقي للمعالجة j ، والمتحولات ε_{ij} تتوزع مستقلة فيما بينها وفق التوزيع الطبيعي ، بمتوسط يساوي الصفر ، وتشتت مشترك يساوي σ^2 . وعلى هذا الأساس لنستعرض الآن ما تصبح علية عبارة مجموع مربعات المعالجات ومجموع مربعات الخطأ فنجد :

$$SSE = \sum_{j=1}^{4} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2$$

$$= \sum_{j=1}^{4} \sum_{i=1}^{n_j} \left[(\mu + \tau_j + \varepsilon_{ij}) - \left(\mu + \tau_j + \frac{\sum_{i=1}^{n_j} \varepsilon_{ij}}{n_j} \right) \right]^2 \quad (21)$$

$$= \sum_{j=1}^{4} \sum_{i=1}^{n_j} \left(\varepsilon_{ij} - \frac{\sum_{i=1}^{n_j} \varepsilon_{ij}}{n_j} \right)^2$$

وبما أن نتائج التجربة هي مجرد عينة من مجتمع كل النتائج الممكنة فيما لو كررنا التجربة بدون توقف ، فمن الطبيعي أن نتساءل عن قيمة التوقع أو ما يسمى بالتوقع الرياضي لمجموع مربعات الخطأ . وبالاستناد إلى الفرض

بأن المتحولات ε_{ij} مستقلة ، وتتوزع وفق التوزيع الطبيعي بمتوسط يساوي الصفر وتشتت مشترك يساوي σ^2 ، يمكن البرهان على العلاقة العامة التالية :

$$E(SSE) = E\left[\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_j)^2\right] \qquad (22)$$
$$= \sum_{j=1}^{k}\sum_{i=1}^{n_j}\left[E\left(\varepsilon_{ij} - \frac{\sum_{i=1}^{n_j}\varepsilon_{ij}}{n_j}\right)^2\right] = \sigma^2\sum_{j=1}^{k}(n_j - 1)$$

وهكذا فإن توقع متوسط مربعات الخطأ هو :

$$E(MSE) = E\left(\frac{SSE}{\sum_{j=1}^{k}(n_j - 1)}\right) = \sigma^2 \qquad (23)$$

وهذا يعني أن متوسط مربعات الخطأ هو تقدير منصف للتشتت σ^2 . وعندما تكون حجوم العينات متساوية أي $n_j = n$ من أجل $j = 1, 2, \dots, k$ تصبح العلاقة (22) كما يلي :

$$E(SSE) = \sum_{j=1}^{k}\sum_{i=1}^{n}\left[E\left(\varepsilon_{ij} - \frac{\varepsilon_{1j} + \varepsilon_{2j} + \dots + \varepsilon_{nj}}{n}\right)^2\right] = k(n-1)\sigma^2 \qquad (24)$$

وبصورة مشابهة نجد من مجموع مربعات المعالجات SST ، وباعتبار $N = \sum_{j=1}^{k} n_j$ ما يلي :

$$SST = \sum_{j=1}^{k} n_j(\bar{x}_j - \bar{x})^2$$

$$= \sum_{j=1}^{k} n_j\left[\left(\mu + \tau_j + \frac{\sum_{i=1}^{n_j}\varepsilon_{ij}}{n_j}\right) - \left(\mu + \frac{\sum_{j=1}^{k} n_j\tau_j}{N} + \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}\varepsilon_{ij}}{N}\right)\right]^2 \qquad (25)$$

$$= \sum_{j=1}^{k} n_j\left[\left(\tau_j - \frac{\sum_{j=1}^{k} n_j\tau_j}{N}\right) + \left(\frac{\sum_{i=1}^{n_j}\varepsilon_{ij}}{n_j} - \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}\varepsilon_{ij}}{N}\right)\right]^2 \qquad (25)$$

وقبل حساب توقع هذه الكمية لا بد أن نقرر النموذج الذي نتبناه أهو النموذج I حيث نعتبر المقادير τ_j أعداداً ثابتة أو النموذج II حيث نعتبر المقادير τ_j متحولات عشوائية تشتتها σ_τ^2 . وسنستعرض كلي الحالتين لنرى أين تقع نقاط الخلاف في التحليل :

النموذج I :

لنفرض أن :

$$\sum_{j=1}^{k} n_j \tau_j = 0 \qquad (26)$$

وهذا الشرط يقابله في حالة عينات متساوية كما رأينا الشرط: $\sum_{j=1}^{k} \tau_j = 0$

الناتج عن تعريف τ_j بالذات . ويمكن أن نبين في هذه الحالة أن :

$$E(SST) = \sum_{j=1}^{k} n_j \tau_j^2 + (k-1)\sigma^2$$

ومنه يكون توقع متوسط مربعات المعالجات MST :

$$E(MST) = E\left(\frac{SST}{k-1}\right) = \sigma^2 + \frac{1}{k-1}\sum_{j=1}^{k} n_j \tau_j^2 \qquad (27)$$

وتصبح هذه النتيجة في حالة عينات متساوية الحجم : $E(MST) =$

$$= \sigma^2 + \frac{n}{k-1}\sum_{j=1}^{k} \tau_j^2$$

النموذج II :

إذا اعتبرنا τ_j متحولات مستقلة فيما بينها ويتوزع كل منها وفق التوزيع الطبيعي بمتوسط يساوي الصفر وتشتت σ_τ^2 ، فعندئذ يمكن برهان النتيجة التالية :

$$E(SST) = (k-1)\sigma^2 + \left[N - \frac{\sum_{j=1}^{k} n_j^2}{N}\right]\sigma_\tau^2 \qquad (28)$$

وقد فرضنا هنا أن المتحولات τ_j مستقلة عن المتحولات ε_{ij} من أجل جميع قيم i و j . وهكذا نجد :

$$E(MST) = E\left(\frac{SST}{k-1}\right) = \sigma^2 + n_0 \sigma_\tau^2 \qquad (29)$$

حيث :

$$n_0 = \frac{1}{k-1}\left(N - \sum_{j=1}^{k} n_j^2 / N\right) \qquad (30)$$

وتصبح هذه النتائج في حالة عينات متساوية الحجم : $n_{\circ} = n$

$$E(MST) = \sigma^2 + n\sigma_\tau^2 \qquad (31)$$

ونعيد الآن كتابة الجدول (١١- ٦) مع إضافة عمودين أحدهما لتوقع متوسط المربعات في حالة النموذج I والآخر لتوقع متوسط المربعات في حالة النموذج II . ونحصل بذلك على الجدول (١١-١٢) .

جدول ١١–١٢ تحليل التشتت في حالة عينات غير متساوية الحجم

مصدر التشتت	درجات الحرية	مجموع المربعات	متوسط المربعات	F	توقع متوسط المربعات النموذج I	توقع متوسط المربعات النموذج II
ما بين المعالجات	$k-1$	SST	$MST=\frac{SST}{k-1}$		$\sigma^2+\frac{1}{k-1}\sum_{j=1}^{k} n_j \tau_j^2$	$\sigma^2+n_0\sigma_\tau^2$
الخطأ	$\sum_{j=1}^{k}(n_j-1)$	SSE	$MSE=\frac{SSE}{\sum_{j=1}^{k}(n_j-1)}$	$\frac{MST}{MSE}$	σ^2	σ^2
المجموع	$\sum_{j=1}^{k} n_j-1$	SS				

والجدول ١١ ـ ٨ المتعلق بالمثال ٨ يصبح على الشكل التالي علماً أن $n_0=4.57$:

جدول ١١ـ١٣ تحليل التشتت للتجربة في المثال ٨

مصدر التشتت	درجات الحرية	مجموع المربعات	متوسط المربعات	F	توقع متوسط المربعات النموذج	توقع متوسط المربعات النموذج
ما بين الأسمدة	3	432	144		$\sigma^2+\frac{1}{3}\sum_{j=1}^{4} n_j \tau_j^2$	$\sigma^2+4.57\,\sigma_\tau^2$
الخطأ	16	32	2	72	σ^2	σ^2
المجموع	19	464				

ولنذكر أننا في كلي المثالين ٧ و ٨ افترضنا أن للمجتمعات الـ k التي نقارنها $k=5$ في المثال ٧ و $k = 4$ في المثال ٨) نفس التشتت σ^2 وهو ما سميناه بخاصة التجانس . وهذه الخاصة هي التي تمكننا من تركيب تشتتات العينات الـ k وهي $S_1^2, \ldots S_k^2$ للحصول على تقدير مركب S^2 للتشتت المشترك σ^2 . هذا التقدير المركب S^2 هو الذي دعيناه في جدول تحليل التشتت بمتوسط مربعات الخطأ MSE

١١ـ٨ إختبار الفرضيات في التصميم التام العشوائية : نعرض الفرضيات هنا بدقة أي بعبارات رياضية يمليها النموذج الرياضي الذي فرضناه . وهكذا نكتب الفرضية في المثال ٨ على الشكل :

$$H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$$

وذلك من أجل النموذج I . وهذا يكافئ كتابة :

$$H_0 : n_1\tau_1^2 + n_2\tau_2^2 + n_3\tau_3^2 + n_4\tau_4^2 = 0$$

وبالطبع فإن الشرط اللازم والكافي ليكون $\sum_{j=1}^{4} n_j \tau_j^2 = 0$ هو أن يكون كل من $\tau_1, \tau_2, \tau_3, \tau_4$ مساوياً للصفر ، ومن أجل النموذج II نكتب الفرضية على الشكل :

$$H_0 : \sigma_\tau^2 = 0$$

وتعني هذه الفرضية أنه لا توجد فروق بين متوسط تأثيرات معالجات الأسمدة الأربعة على إنتاج القمح . أو بعبارة أخرى نقول أن متوسطات المجتمعات الأربعة متساوية ، ذلك لأن :

$$E(x_{ij}) = E(\mu + \tau_j + \varepsilon_{ij}) = \mu + \tau_j \quad , \quad \begin{matrix} i = 1, \ldots, n_j \\ j = 1, \ldots, 4 \end{matrix}$$

وإذا كانت الفرضية H_0 صحيحة فإن $E(x_{ij}) = \xi_j = \mu$ حيث ξ_j هو متوسط

المجتمع الموافق للمعالجة أ ، وبالتالي فإن $\xi_1 = \xi_2 = \xi_3 = \xi_4 = \mu$ وهو الشكل الذي عرضناه ، في البداية ، للفرضية التي تهدف طريقة تحليل التشتت لإختبارها .

ويؤدي النموذجان I و II إلى نفس الإختبار لِ H_0 في مثالناهنا ، ولذلك سنناقشهما معاً . وسوف لا يكون الأمر كذلك دائماً على أي حال .

وبالعودة إلى الجدول (١١-١٣) نجد أن متوسط مربعات « ما بين الأسمدة » هو تقدير منصف للمقدار $\sigma^2 + \frac{1}{3}\sum_{j=1}^{4} n_j \tau_j^2$ (تحت النموذج I)، أو تقدير منصف لِ $\sigma^2 + 4.57\sigma_\tau^2$ (تحت النموذج II) ، بينما يشكل متوسط مربعات الخطأ MSE تقديراً منصفاً لِ σ^2 . ويمكن البرهان على أن توزيع الكمية :

$$\frac{SSE}{\sigma^2} = \frac{\sum_{j=1}^{k} (n_j - 1) s^2}{\sigma^2} \qquad (32)$$

هو التوزيع χ^2 بِـ $\sum_{j=1}^{k}(n_j-1)$ درجة من الحرية . وهذه الظاهرة ، وهي صحيحة بصرف النظر عن الفرضية التي نضعها حول تأثيرات المعالجات ، تشجّع على تأمل النسبة المقابلة من أجل مجموع مربعات « ما بين المعالجات » أي النسبة :

$$\frac{SST}{\sigma^2 + \frac{1}{3}\sum_{j=1}^{4} n_j \tau_j^2} \qquad (33)$$

أو

$$\frac{SST}{\sigma^2 + n_0 \sigma_\tau^2} \qquad (34)$$

وفقاً لحالة النموذج I أو النموذج II على الترتيب . ونُبرهن في الإحصاء

النظري أن الكميتين SST و SSE مستقلتان . لنشكل الآن النسبتين :

$$\frac{SST/(\sigma^2 + \sum_{j=1}^{k} \frac{n_j \tau_j^2}{k-1})(k-1)}{SSE/\sigma^2 \sum_{j=1}^{k} (n_j - 1)} \qquad (35)$$

$$\frac{SST/(\sigma^2 + n_0 \sigma_\tau^2)(k-1)}{SSE/\sigma^2 \sum_{j=1}^{k} (n_j - 1)} \qquad (36)$$

وفقاً لحالتي النموذج I أو النموذج II على الترتيب . ومن سوء الحظ فإننا لا نعلم قيم σ^2 ، σ_τ^2 أو τ_j $(j = 1, 2, ..., k)$. ولكن نلاحظ أنه اذا كانت الفرضية الإبتدائية صحيحة فإن كلاً من العلاقتين (35) و (36) تصبحان على الشكل :

$$F = \frac{SST/\sigma^2(k-1)}{SSE/\sigma^2 \sum_{j=1}^{k}(n_j-1)}$$

$$= \frac{SST/(k-1)}{SSE/\sum_{j=1}^{k}(n_j-1)} = \frac{MST}{MSE} \qquad (37)$$

ويمكن البرهان على أن النسبة SST/σ^2 تتوزع ، علماً أن الفرضية الإبتدائية صحيحة ، وفق التوزيع χ^2 بـ$(k-1)$ درجة من الحرية . وكما نذكر من الفقرة (٦ – ٦) فإن توزيع النسبة F في (37) هو التوزيع $F(\nu_1, \nu_2)$ حيث $\nu_1 = k-1$ و $\nu_2 = \sum_{j=1}^{k}(n_j-1)$ (وفي حالة تساوي العينات يكون $\nu_1 = k-1$ ، $\nu_2 = k(n-1)$) . وهكذا نستنتج أنه اذا كانت الفرضية الابتدائية صحيحة فإن توزيع النسبة F في تحليل التشتت هو توزيع سنيديكور أو التوزيع F وبالتالي يمكن استخدام هذه النسبة كإحصاء لإختبار الفرضية الإبتدائية H_0 .

ونلاحظ أن النسبة F في كلي المثالين ٧ و ٨ كانت بحيث رفضنا الفرضية

H_0 . وبالطبع سيواجه المجرب حالات تكون فيها قيمة F بحيث لا نستطيع رفض H_0 . ولكن هب أننا حصلنا على قيمة لِـ F أصغر من الواحد فهل نكتفي بالقول بأننا لا نستطيع رفض الفرضية H_0 أي نقبلها ؟ وفي الحقيقة فإن الاكتفاء بمثل هذا القول يتجاهل حالة هامة جداً . وقد نقول في مثل هذه الحالة أن التغيرات التي تنشأ من عينة إلى أخرى هي التي سببت المبالغة في تقدير توقع المخرج (أي تقدير s^2)، وأنها على العكس بخست تقدير توقع الصورة حقه . ومع كون مثل هذا التعليل منطقياً ، فماذا نقول لو أن قيمة النسبة F كانت صغيرة إلى الحد الذي يصبح معه مقلوبها $F' = \frac{1}{F}$ أكبر من $F_{1-\alpha}(\nu_2, \nu_1)$ ؟ وبما أن مقلوب توزيع $F(\nu_1, \nu_2)$ هو التوزيع $F(\nu_2, \nu_1)$ فإنه يبدو وكأننا يجب أن نرفض شيئاً ما . ويبدو معقولاً أن نقول في مثل هذه الحالة أنه يجب أن نرفض النموذج الرياضي الذي اعتمدته التجربة ونبحث عن نموذج يتلاءم بشكل أفضل مع الظاهرة المدروسة .

١١ ـ ٩ إختبار درجات الحرية كل بمفردها : اذا احتوت التجربة k من المعالجات فيمكن تقسيم « مجموع مربعات المعالجات » إلى $k-1$ من الأجزاء كل منها مجموع مربعات ثم إختبار كل منها بمقارنتها مع متوسط مربعات الخطأ .

ولتوضيح الفكرة لنعد إلى البيان الإحصائي في المثال ٨ . ولنفرض أننا لاحظنا عند تخطيط التجربة وجود شبه كبير في تركيب كل من السمادين رقم 1 ورقم 2 ، ووجود شبه كبير أيضاً بين تركيبي السمادين رقم 3 ورقم 4 ، ولكن السمادين رقم 1 و 2 يختلفان اختلافاً بيّناً عن السمادين 3 و 4 . فيبدو من المنطقي أن نقيم مقارنة بين : (i) المعالجتين 1 و 2 في مقابل المعالجتين 3 و 4 (ii) المعالجة 1 في مقابل المعالجة 2 و (iii) المعالجة 3 في مقابل المعالجة 4 .

ونعرّف جبرياً المقارنة بين الكميات $W_1, \dots, W_n$ (حيث w_i هو مجموع

r_i من الملاحظات) كما يلي :

$$c_i = c_{i1} w_1 + c_{i2} w_2 + \ldots + c_{in} w_n \qquad (38)$$

حيث :

$$\sum_{j=1}^{n} r_j c_{ij} = 0 \qquad (39)$$

وإذا كان كل r_i مساوياً لِـ r أي كانت المقادير $w_1, \ldots, w_n$ مجاميعاً لأعداد متساوية من الملاحظات ، فعندئذ يصبح الشرط الضروري لمقارنة على الشكل :

$$\sum_{j=1}^{n} c_{ij} = 0 \qquad (40)$$

ويمكن تمثيل المقارنات الثلاث بين المعالجات التي ذكرناها أعلاه كما يلي :

$$C_i = \sum_{j=1}^{4} c_{ij} T_j$$

حيث :

$$\begin{aligned} C_1 &= (7)(184) + (7)(68) + (-3)(170) + (-3)(378) \\ C_2 &= (1)(184) + (-2)(68) + (0)(170) + (0)(378) \\ C_3 &= (0)(184) + (0)(68) + (9)(170) + (-5)(378), \end{aligned} \qquad (41)$$

و T_j ترمز كالمعتاد لمجموع المعالجة j .

ونلاحظ أن الأمثال هنا تحقق الشرط المطلوب ($r_1 = 4$ ، $r_2 = 2$ ، $r_3 = 5$ و $r_4 = 9$ كما نرى من البيان الإحصائي في الجدول ١١ـ٧) في العلاقة (39) .

ويمكن أن نتساءل هنا عن كيفية الحصول على الأمثال c_{ij} $(i = 1,2,3; j = 1,2,3,4)$ المستخدمة في المقارنات (41). ولنأخذ على سبيل المثال المقارنة C_1 وهي مقارنة بين متوسط ست ملاحظات $(4+2)$ مع متوسط 14 ملاحظة $(9+5)$. وبما أن المضاعف المشترك البسيط لِـ 6 و 14 هو 42 ،

فإن التثقيلات الضرورية لمقارنة مجاميع أعداد غير متساوية من الملاحظات في المقارنة C_1 هي 7 و 3 . ويمكن إيجاد الأمثال في المقارنات الأخرى بنفس الطريقة .

لنحسب الآن مجاميع المربعات الموافقة للمقارنات الثلاث C_1 ، C_2 و C_3 بين المعالجات المذكورة في المثال ٨ فنجد :

$$C_1 \text{ مجموع مربعات المقارنة } = \frac{(T_1 + T_2)^2}{n_1 + n_2} + \frac{(T_3 + T_4)^2}{n_3 + n_4} - \frac{T^2}{\sum_{j=1}^{4} n_j}$$

$$= \frac{(184 + 68)^2}{6} + \frac{(170 + 378)^2}{14} - \frac{(800)^2}{20} = 34.3$$

$$C_2 \text{ مجموع مربعات المقارنة } = \frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} - \frac{(T_1 + T_2)^2}{n_1 + n_2}$$

$$= \frac{(184)^2}{4} + \frac{(68)^2}{2} - \frac{(252)^2}{6} = 192.0$$

$$C_3 \text{ مجموع مربعات المقارنة } = \frac{T_3^2}{n_3} + \frac{T_4^2}{n_4} - \frac{(T_3 + T_4)^2}{n_3 + n_4}$$

$$= \frac{(170)^2}{5} + \frac{(378)^2}{9} - \frac{(584)^2}{14} = 205.7$$

ونلاحظ أن $432 = 205.7 + 192.0 + 34.3$ وهو مجموع مربعات ما بين الأسمدة في الجدول (١١ ـ ٨) . ونلاحظ أنه كان يمكن حساب مجاميع المربعات هذه بالاستفادة من العلاقة :

$$\text{مجموع مربعات المقارنة} = \frac{C_i^2}{\sum_{j=1}^{4} n_j c_{ij}^2} \qquad (42)$$

ونلاحظ في المقارنات الثلاث C_1 C_2 و C_3 أنه لا يتحقق الشرط :

$$\sum_{j=1}^{n} r_i c_{ij} = 0 \qquad \text{(من أجل أي j)} \qquad (43)$$

فقط وإنما يتحقق أيضاً الشرط :

$$\sum_{j=1}^{n} r_j c_{ij} c_{kj} = 0 \quad (i \neq k) \qquad (44)$$

ونقول في هذه الحالة أن المقارنة C_i متعامدة مع المقارنة C_k.

وفي حالة عينات من نفس الحجم تكون $r_j = r$ من أجل جميع قيم j وتبقى العلاقات السابقة صحيحة مع التعديل الضروري للحالة الجديدة . ويصبح مثلاً شرط التعامد :

$$\sum_{j=1}^{n} c_{ij} c_{kj} = 0 \quad (i \neq k) \qquad (45)$$

ولو لم تكن المقارنات الثلاث أعلاه متعامدة مثنى مثنى لما كان مجموع مجاميع مربعاتها مساوياً لمجموع مربعات المعالجات . ونؤكد على ضرورة صياغة المقارنات التي نريد اختبارها مقدماً وقبل تحليل البيان الإحصائي .

وفي المثال ٨ يمكن كتابة تحليل التشتت كما في الجدول (١١–١٤) . ونقوم بالإختبار F لكل مقارنة بالطريقة المعتادة . ونلاحظ أننا نرفض الفرضية الابتدائية في كل من المقارنات الثلاث . أي أن نتائج الإختبار F تشير إلى وجود فروق هامة ضمن كل من المقارنات الثلاث ، وهذا ما نعبر عنه بقولنا أن المقارنات الثلاث هامة . وربما كان هذا نتيجة لكون المعلومات الإحصائية التي نحللها معلومات مصطنعة . وكثيراً ما تسهم عملياً مقارنة أو أكثر في تشكيل مجموع مربعات المعالجات بينما يبقى ما تقدمه مقارنات أخرى صغيراً نسبياً ، مما يشير في مثل هذه الحالة الى المصادر الأكثر أهمية التي تسبب وجود فروق هامة بين المعالجات .

جدول ١١ـ١٤ تحليل تشتت المقارنات الثلاث المتعلقة بالبيان الإحصائي في الجدول ١١ـ٧

مصدر التغير	درجات الحرية	مجموع المربعات	متوسط المربعات	توقع متوسط المربعات
المعالجتان 1 و 2 مقابل المعالجتين 3 و 4	1	34.3	43.3	$\sigma^2 + \frac{[(n_3+n_4)(n_1\tau_1+n_2\tau_2)-(n_1+n_2)(n_3\tau_3+n_4\tau_4)]^2}{(n_1+n_2)(n_3+n_4)\sum_{j=1}^{4} n_j}$
المعالجة 1 مقابل المعالجة 2	1	192.0	192	$\sigma^2 + \frac{n_1 n_2}{n_1+n_2}(\tau_1-\tau_2)^2$
المعالجة 3 مقابل المعالجة 4	1	205.7	205.7	$\sigma^2 + \frac{n_3 n_4}{n_3+n_4}(\tau_3-\tau_4)^2$
الخطأ	16	32	2	σ^2
المجموع	19	464		

لنأخذ الآن مثالاً عن حالة عينات متساوية الحجم . فنعود إلى المثال ٧ ، والبيان الإحصائي في الجدول ١١ـ٧ . ولنفرض أننا نهتم بالمقارنات المبينة في الجدول (١١ـ١٥) . وأننا قررنا هذه المقارنات سلفاً أي قبل الإطلاع على نتائج التجربة .

جدول ١١ـ١٥ مقارنات بين المعالجات في المثال ٧

المقارنات	الأسمدة 1	2	3	4	5
C_1	−1	+4	−1	−1	−1
C_2	+1	0	+1	−1	−1
C_3	+1	0	−1	0	0
C_4	0	0	0	+1	−1

ويمكن أن نبدأ هنا بنفس الطريقة التي حللنا فيها المثال السابق . ولكن الحسابات هنا ستكون أسرع بكثير . فالمقارنة C_1 مثلاً هي :

$$(-1)\ (180) + (4)\ (160) + (-1)\ (160) + (-1)\ (164) + (-1)\ (136) = 0$$

ومجموع المربعات العائدة للمقارنة C_1 هو :

مجموع مربعات المقارنة $C_1 = \frac{(0)^2}{20(4)} = 0$

وبصورة مماثلة نجد أن :

مجموع مربعات $C_2 = \frac{(40)^2}{4(4)} = 100$

وهكذا . وبصورة عامة لدينا :

$$\overline{\overline{C}}_i = \sum_{j=1}^{n} c_{ij}\, w_i \qquad (46)$$

حيث w_j هي مجموع r من الملاحظات . ومجموع المربعات معطى ، بصورة عامة ، بالعلاقة :

$$\frac{C_i^2}{r\sum_{j=1}^{n} c_{ij}^2} = \frac{\left(\sum_{j=1}^{n} c_{ij} w_j\right)^2}{r\sum_{j=1}^{n} c_{ij}^2} \qquad (47)$$

وهي حالة خاصة من المعادلة **(42)** تكون فيها $r_j = r$ مهما كانت j . ونحصل هنا على تحليل التشتت المبين في الجدول (١١–١٦) .

ونلاحظ في هذا المثال أن ثلاثاً من المقارنات الأربعة تشكل كامل مجموع مربعات ما بين المعالجات ، وأن المقارنة الباقية ، وقد وُضعت قبل الاطلاع على نتائج التجربة أو قبل تنفيذها ، لم تقدم شيئاً البتة . ويجب ألا نفترض هنا أنه يمكن اللجؤ إلى الاختبار t لمقارنة ثنائية بين المعالجات بعد أن يكون الاختبار F قد أشار إلى وجود مثل هذه الفروق . ومثل هذا العمل ليس مشروعاً ، بصورة عامة . فمع أن التوزيع $F_{(1,\nu)}$ يكافيء ، كما رأينا في الفقرة (٦ – ٦) ، التوزيع $t^2_{(\nu)}$ ، وأن الإختبار F المتعلق بالمقارنات هو عملياً مكافئ للإختبار t ، إلا أن المقارنات هنا قد **صيغت** سلفاً قبل تنفيذ التجربة وليس بعد الإطلاع على نتائجها أو بعد تحليل التجربة والخروج بنتائج محددة .

جدول ١١ـ١٦ تحليل التشتت لمقارنات تتعلق بالبيان الإحصائي في الجدول ١١ـ٤

مصدر التغير	درجات الحرية	مجموع المربعات	متوسط المربعات	توقع متوسط المربعات
المعالجة 2 في مقابل بقية المعالجات	1	0	0	$\sigma^2 + \frac{4}{20}(\tau_1+\tau_3+\tau_4+\tau_5-4\tau_2)^2$
المعالجة 1 و 3 مقابل المعالجتين 4 و 5	1	100	100	$\sigma^2 + \frac{4}{4}(\tau_1+\tau_3-\tau_4-\tau_5)^2$
المعالجة 1 مقابل المعالجة 3	1	50	50	$\sigma^2 + \frac{4}{2}(\tau_1-\tau_3)^2$
المعالجة 4 مقابل المعالجة 5	1	98	98	$\sigma^2 + \frac{4}{2}(\tau_4-\tau_5)^2$
الخطأ	15	130	8.67	σ^2
المجموع	19	378		

١١-١٠ الفرق المهم الأدنى والمقارنات بدرجة واحدة من الحرية :

بقيت مسألة كيفية القيام بمقارنات معينة بين متوسطات المعالجات تواجه الباحثين العلميين والإحصائيين لسنين عديدة . وقد تطرقنا في الفقرة الماضية إلى القيام بمثل هذه المقارنات عندما تتم صياغتها سلفاً وقبل تنفيذ التجربة أو الإطلاع على نتائجها . ولكن معرفة المقارنات الهامة الضرورية بصورة سلفية غير ممكن دائماً . فالتجربة غالباً ما تكون من طبيعة استكشافية ، وبالتالي فإننا ننتظر من نتائج التجربة ومن التحليل الإحصائي لهذه النتائج أن تقدم لنا شيئاً من المعرفة عن طبيعة المعالجات المدروسة ومواقعها من بعضها البعض ، وبالتالي فإن النتائج هي التي قد تشير إلى المقارنات الهامة ، أو أنها تثير تساؤلات جديدة لها أهميتها تستدعي القيام بمقارنات معينة بين المعالجات . ومن المؤكد أن المجرب يرغب في أن يحصل من بيانه الإحصائي على أكثر من مجرد عبارة بسيطة (مبنية على الإختبار F) تقول ، مثلاً ، أن التأثيرات الحقيقية للمعالجات المدروسة ليست جميعها متساوية . فهو يريد ، بعد أن يصل إلى هذه النتيجة ، أن يعرف أيضاً ما إذا كان يمكن إعتبار عدد من هذه المعالجات مكافئة لبعضها البعض ، ومعرفة تلك المعالجة من المعالجات المدروسة التي يمكن اعتبارها المعالجة « الأفضل » .

ويقترح سنيديكور الحصول على تقدير مجالي (أي إقامة مجال ثقة) بالنسبة للمتوسط الحقيقي لكل معالجة ثم ملاحظة المجالات التي تتقاطع مع بعضها . وبعض الباحثين يفضلون استخدام ما يسمى بالفرق المهم الأدنى ، وسنرمز اختصاراً بِـ (ف.م.أ) ، وهو معرف بالعبارة التالية :

$$\text{ف.م.أ} = t_{\alpha} \cdot S_{(\bar{x}_i - \bar{x}_j)} \qquad (48)$$

حيث α هو مستوى الأهمية الذي اخترناه للتجربة ، ودرجات الحرية لِ t هو عدد درجات الحرية الموافقة « للخطأ » في جدول تحليل التشتت .
والطريقة عندئذ هي أن نأخذ القيمة المطلقة للفرق بين متوسطي معالجتين ، مثلاً $|\bar{x}_1 - \bar{x}_2|$، ونقول بأن التأثيرين الحقيقيين لهاتين المعالجتين غير متساويين اذا تجاوز هذا المقدار الفرق المهم الأدنى (ف.م.أ) المعرف في العلاقة(48).
أي اذا كان :

$$|\bar{x}_1 - \bar{x}_2| > t_{\alpha} \cdot S_{(\bar{x}_1 - \bar{x}_2)}$$

وبصورة عامة يُعتبر تطبيق هذه القاعدة أمراً خطراً اذا استُخدمت بدون تمييز بالنسبة لكل الأزواج الممكنة من المعالجات . ويبدو من المنطقي في حال وجود « معالجة أساس » ، وهي معالجة معيارية ، أن نقول بأن تطبيق طريقة الفرق المهم الأدنى يصبح مشروعاً وذلك لمقارنة بقية المعالجات مع هذه المعالجة المعيارية .

نلاحظ في طريقة تحليل التشتت أننا قسمنا مجموع المربعات الكلي إلى قسمين أحدهما يتضمن تشتت المتوسطات الملحوظة للمعالجات والآخر يحوي ما سميناه بمجموع مربعات الخطأ . واستخدمنا هذين الجزئين لإختبار فرضية تساوي متوسطات المعالجات . فإذا كانت المتوسطات الملحوظة قريبة من بعضها البعض نقبل الفرضية أما إذا كانت مبعثرة بصورة ملحوظة نرفض الفرضية . وكما نعلم فقد استخدمنا تشتت المتوسطات الملحوظة $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$ لقياس مدى تبعثرها .

ولكن المدى w هو أيضاً قياس للتبعثر ويمكن استخدامه لنقيس تبعثر هذه المتوسطات ويمكن إختبار فرضية تساوي متوسطات المعالجات بمقارنة مدى المتوسطات الملحوظة مع مجموع مربعات الخطأ . ويمكن القيام بذلك

باستخدام إحصاء الإختبار :

$$q = \frac{w}{S_p/\sqrt{n}}$$

حيث S_p^2 هو متوسط مربعات الخطأ و n حجم كل من العينات الـ k التي تحويها التجربة . والمخرج $S_p/\sqrt{n}$ هو تقدير لـ $\sigma/\sqrt{n}$ (الإنحراف المعياري للمتوسطات $\bar{x}$ لعينات حجمها n مسحوبة من نفس المجتمع ، وحيث σ هو الإنحراف المعياري لهذا المجتمع) .

ويقدم جدول في الملحق عدة نسب مئوية لتوزيع الإحصاء q حيث k عدد المتوسطات و ν عدد درجات الحرية الخاص بالخطأ في جدول تحليل التشتت . وقد كُتب في رأس الجدول $q = w/s$ وهي النسبة الموافقة للحالة n = 1 أي حالة k من الملاحظات من مجتمع طبيعي . ويُعدّل الإحصاء بالعامل $\sqrt{n}$ عند تطبيقه في حالة متوسطات لعينات حجم كل منها n .

ولنفرض أن لدينا أربع ملاحظات (n = 4) من كل من ثلاثة مجتمعات k = 3 ، فعدد درجات الحرية الموافق لمتوسط مربعات الخطأ هو 9 = 3 – 12. ونقرأ في الجدول أن احتمال كون $q = \frac{w}{S_p/4}$ أقل من 3.95 هو 0.95 وأن احتمال كونه أقل من 5.43 هو 0.99 . ومنطقة الرفض الموافقة ، عند مستوى الأهمية $\alpha = .05$ ، لإختبار الفرضية القائلة بأن متوسطات المجتمعات الثلاثة متساوية ، هذه المنطقة تتألف من قيم إحصاء الأختبار q الأكبر من 3.95 وتكون هذه المنطقة ، من أجل اختبار عند مستوى الأهمية $\alpha = .01$ ، هي قيم q الأكبر من 5.43 . وإذا فرضنا ، على سبيل المثال ، أن المتوسطات الملحوظة هي $\bar{x}_1 = 2.25$ ، $\bar{x}_2 = 4.00$ ، $\bar{x}_3 = 4.50$ وأن متوسط مربعات الخطأ $S_p^2 = 4.41$ ، فعندئذ تصبح القيمة الملحوظة لإحصاء الإختبار q هي $q = \frac{2.25}{(2.10)/2} = 2.14$ ، وهي أقل من 3.95 . وهكذا فإننا لا نرفض الفرضية عند المستوى $\alpha = 0.5$.

تعميم إلى فرضيات أخرى : يمكن الاستفادة من توزيع q من أجل التقدير ، والقيام بإختبار عدة فرضيات بنفس الوقت حول فروق ممكنة بين المتوسطات أي حول مقادير من النوع $\mu_4-\mu_2$ ، $\mu_1-\mu_2$ ، $\mu_1-\mu_3$.

وبصورة عامة حول أي مقارنة بين المتوسطات الـ k مثل $a_1\mu_1+a_2\mu_2+\cdots+a_k\mu_k$ حيث $\sum_{i=1}^{k} a_i=0$ وحيث نأخذ للسهولة مجموع المقادير a الموجبة مساوياً للواحد.

وتعتمد الإختبارات المتعلقة بالمقارنات هذه على النظرية التالية :

من أجل عينات عشوائية من k من المجتمعات الطبيعية لها نفس التشتت ، يكون احتمال أن تحقق كل المقارنات ، وبنفس الوقت ، المتراجحة :

$$-\frac{q_{1-\alpha}\cdot S_p}{\sqrt{n}} < (a_1\bar{x}_1+\cdots+a_k\bar{x}_k)-(a_1\mu_1+\cdots+a_k\mu_k) < \frac{q_{1-\alpha}\cdot S_p}{\sqrt{n}} \quad (49)$$

مساوياً لِ $1-\alpha$ حيث $q_{1-\alpha}$ هي قيمة q الموافقة كما نجدها في الجدول.

وعلى سبيل المثال ، إذا اعتبرنا الحالة التي ناقشناها أعلاه والتي وجدنا فيها $k=3$ ، $n=4$ ، $\bar{x}_1=2.25$ ، $\bar{x}_2=4.00$ $\bar{x}_3=4.50$ $S_p^2=4.41$ ومن أجل أمثال ثقة تساوي 0.95 نستخدم $q_{.95}=3.95$ كما نجدها من الجدول من أجل $k=3$ وعدد من درجات الحرية $\nu=9$ (وباستخدام الرمز $q_{1-\alpha}(k,\nu)$ نكتب $q_{.95}(3,9)=3.95$). نحسب الآن $q_{1-\alpha}\cdot S_p/\sqrt{n}=3.95(2.10)/\sqrt{4}=4.15$

وعندها يمكننا القول ، وبثقة تساوي 0.95 ، أن العبارات التالية كلها صحيحة بنفس الوقت :

$$\bar{x}_1-\bar{x}_2-4.15<\mu_1-\mu_2<\bar{x}_1-\bar{x}_2+4.15$$

$$-5.90<\mu_1-\mu_2<2.40$$

و

$$\bar{x}_1-\bar{x}_3-4.15<\mu_1-\mu_3<\bar{x}_1-\bar{x}_3+4.15$$

$$-6.40<\mu_1-\mu_3<1.90$$

$$\bar{X}_1 - \bar{X}_3 - 4.15 < \mu_2 - \mu_3 < \bar{X}_2 - \bar{X}_3 + 4.15$$ و ×

$$-4.65 < \mu_2 - \mu_3 < 3.65$$

و ×

$$\frac{\bar{X}_1 + \bar{X}_2}{2} - \bar{X}_3 - 4.15 < \frac{\mu_1 + \mu_2}{2} - \mu_3 < \frac{\bar{X}_1 + \bar{X}_2}{2} - \bar{X}_3 + 4.15$$

$$-5.53 < \frac{\mu_1 + \mu_2}{2} - \mu_3 < 2.77$$

و

$$\frac{\bar{X}_1 + \bar{X}_3}{2} - \bar{X}_2 - 4.15 < \frac{\mu_1 + \mu_3}{2} - \mu_2 < \frac{\bar{X}_1 + \bar{X}_3}{2} - \bar{X}_2 + 4.15$$

$$-4.77 < \frac{\mu_1 + \mu_3}{2} - \mu_2 < 3.53$$

و

$$\frac{\bar{X}_2 + \bar{X}_3}{2} - \bar{X}_1 - 4.15 < \frac{\mu_2 + \mu_3}{2} - \mu_1 < \frac{\bar{X}_2 + \bar{X}_3}{2} - \bar{X}_1 + 4.15$$

$$-2.15 < \frac{\mu_2 + \mu_3}{2} - \mu_1 < 6.15$$

وهكذا من أجل أية مجموعة من المقارنات نستطيع أو نرغب في كتابتها .

وهذه الطريقة مفيدة جداً باعتبارها تمكننا من كتابة مجالات ثقة حول فروق أو مقارنات يقترحها البيان الإحصائي نفسه بدلاً من الإقتصار على مقارنات خاصة نضعها سلفاً قبل الإطلاع على نتائج التجربة . ونستخدم النتائج الاستكشافية لتجربة ، عادة ، من أجل صياغة فرضية جديدة تكون موضوع تجربة جديدة نقوم بها للوصول إلى نتيجة برفض أو قبول هذه الفرضية الجديدة . ولكن باستخدام الإحصاء q يمكننا الإستفادة من نفس البيان الإحصائي لإختبار فرضيات تمليها نتائج غير متوقعة تمخضت عنها التجربة . ولا بد من التأكيد في هذا المجال أنه إذا كان لمقارنة معينة بمفردها أهمية

خاصة فإنه يمكن الحصول على مجال ثقة أقصر حول هذه المقارنة باستخدام التوزيع t وكتابة مجال ثقة من النوع :

$$t_{\frac{1}{2}\alpha} \cdot S_p \sqrt{\frac{\sum a_i^2}{n}} < (a_1 \bar{x}_1 + \cdots + a_k \bar{x}_k) - (a_1 \mu_1 + \cdots + a_k \mu_k) < t_{1-\frac{1}{2}\alpha} \cdot S_p \sqrt{\frac{\sum a_i^2}{n}}$$

إلا أن أمثال الثقة $(1-\alpha)$ ينطبق فقط على هذه المقارنة بمفردها وليس على جملة من المقارنات . أي أننا نستطيع القول قبل تنفيذ التجربة بأن احتمال أن تكون هذه العبارة صحيحة هو $(1-\alpha)$ ، ولكننا لا نستطيع تعميم هذا الاحتمال على عبارات أخرى بنفس الوقت ، أي استخدام t لوضع مجالات ثقة مشابهة لمقارنات أخرى والقول بأن احتمال أن تكون هذه المجالات كلها صحيحة بنفس الوقت لا يزال $(1-\alpha)$. بينما يصبح مثل ذلك ممكناً باستخدام الإحصاء q ، ولكن المجالات الناتجة عن استخدام q ستكون أعرض .

ويمكن تسجيل عدد من المقارنات على شكل جدول من النوع المبين في الجدول (١١- ١٧) حيث وضعنا الأمثال $a_1, a_2, \ldots, a_k$ الموافقة لِ $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$ بالنسبة لكل مقارنة في العمود الأول . ويبين الجدول (١١- ١٨) تطبيقاً للجدول (١١-١٧) في حالة مثال عددي . وفي هذا المثال ، ومستخدمين الإحصاء q عند المستوى α ، لا نجد دلالة بالنسبة للمقارنات الستة تمكننا من رفض القيمة صفر ، باعتبار أن كلاً من المجالات الستة التي حصلنا عليها تحوي القيمة صفر .

ولوفرضنا أننا سنقدر مقارنة واحدة فقط من بين هذه المقارنات لكان يمكن استخدام الإحصاء t من أجل هذه المقارنة الوحيدة . أي لو أننا اخترنا ، مثلاً ، المقارنة الثالثة قبل تنفيذ التجربة فيمكننا عندئذ وضع 95. مجال ثقة من أجل الفرق $\mu_1 - \mu_2$ كما يلي :

$$-.50 \pm 2.26(2.10)\sqrt{\frac{2}{4}} = -.50 \pm 3.36$$

جدول ١١–١٧ أمثلة عن مقارنات بين ثلاث معالجات

المقارنة بين متوسطات المجتمعات الثلاثة	حدود مجال الثقة	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$
$\mu_1 - \mu_2$	$(\bar{x}_1 - \bar{x}_2) \pm \frac{q \cdot S_p}{\sqrt{n}}$	1	−1	0
$\mu_1 - \mu_3$	$(\bar{x}_1 - \bar{x}_3) \pm \frac{q \cdot S_p}{\sqrt{n}}$	1	0	−1
$\mu_2 - \mu_3$	$(\bar{x}_2 - \bar{x}_3) \pm \frac{q \cdot S_p}{\sqrt{n}}$	0	1	−1
$\frac{\mu_1 + \mu_2}{2} - \mu_3$	$\left(\frac{\bar{x}_1 + \bar{x}_2}{2} - \bar{x}_3\right) \pm \frac{q \cdot S_p}{\sqrt{n}}$	$\frac{1}{2}$	$\frac{1}{2}$	−1
$\frac{\mu_1 + \mu_3}{2} - \mu_2$	$\left(\frac{\bar{x}_1 + \bar{x}_3}{2} - \bar{x}_2\right) \pm \frac{q \cdot S_p}{\sqrt{n}}$	$\frac{1}{2}$	−1	$\frac{1}{2}$
$\frac{\mu_2 + \mu_3}{2} - \mu_1$	$\left(\frac{\bar{x}_2 + \bar{x}_3}{2} - \bar{x}_1\right) \pm \frac{q \cdot S_p}{\sqrt{n}}$	−1	$\frac{1}{2}$	$\frac{1}{2}$

جدول ١١ـ١٨ مثال عددي من أجل المقارنات في الجدول ١١ـ١٧

المقارنة بين المتوسطات	حدود الثقة	2.25	4.00	4.50
$\mu_1 - \mu_2$	-1.75 ± 4.15	1	-1	0
$\mu_1 - \mu_3$	-2.25 ± 4.15	1	0	-1
$\mu_2 - \mu_3$	$-.50 \pm 4.15$	0	1	-1
$\frac{\mu_1 + \mu_2}{2} - \mu_3$	-1.38 ± 4.15	$\frac{1}{2}$	$\frac{1}{2}$	-1
$\frac{\mu_1 + \mu_3}{2} - \mu_2$	$-.62 \pm 4.15$	$\frac{1}{2}$	-1	$\frac{1}{2}$
$\frac{\mu_2 + \mu_3}{2} - \mu_1$	2.00 ± 4.15	-1	$\frac{1}{2}$	$\frac{1}{2}$

١١-١١ التصنيف الثنائي بملاحظة واحدة : سنقدم في هذه الفقرة تحليل تجارب مصممة لدراسة مجتمعات مصنفة وفق خاصتين . ويمكن تنفيذ التجربة بطريقة ندرس فيها عدة متحولات بنفس الوقت . ونختار من أجل كل متحول عدداً من الصفات أو المستويات لدراستها .

ففي دراسة إنتاجية سلالات مختلفة من القمح يمكن أن نتقصى بنفس الوقت تأثيرات أسمدة مختلفة على إنتاجية هذه السلالات . وقد يرغب الباحث الإجتماعي ، عند دراسته لحجم الأسرة ، أن يتقصى تأثيرات حجم المدينة والمناطق ضمن البلاد التي تجري فيها الدراسة ، على حجم الأسرة . ويمكن حساب متوسط حجم العائلة في خمسة أصناف من المدن (مصنفة وفقاً لحجمها) وذلك في كل من ست مناطق ، ثم نتقصى النتائج فيما يتعلق بتأثير حجم المدينة على حجم الأسرة مستقلاً عن تأثير المنطقة ، وتأثير المنطقة على حجم الأسرة مستقلاً عن حجم المدينة .

وسندرس في هذه الفقرة الحالة التي نأخذ فيها ملاحظة واحدة من أجل كل تركيب من المستويات . ويمثل الجدول (١١-١٩) وهو يحوي 12 خلية تجربة بمتحولين ، وبقع متحول الصفوف في ثلاث مستويات ، هي ، مثلاً ، a ، b ، c ويقع متحول الأعمدة في أربع مستويات ، مثلاً ، A ، B ، C ، D . والملاحظة الواقعة في الخلية ij ، أي الخلية الموافقة للصف i والعمود j ، نرمز لها بـ x_{ij} ، أما مجموع الصف i فنرمز له بـ $T_{i.}$ ، أي أن $T_{i.}=\sum_{j=1}^{4} x_{ij}$ ، ومجموع العمود j نرمز له بـ $T_{.j}$. وتمثل النقطة بصورة عامة عملية جمع أو أخذ متوسط فوق جميع الملاحظات الموافقة لقيم الدليل الذي حلّت النقطة محله . وترمز $T_{..}$ للمجموع الكلي للملاحظات . وهكذا يكون $\bar{x}_{i.}$ متوسط الصف i ، و $\bar{x}_{.j}$ متوسط العمود j ، $\bar{x}_{..}$ المتوسط الإجمالي لكافة الملاحظات . وبصورة عامة نتحول i من 1 إلى r و j من 1 إلى c ، أي يوجد r من الصفوف و c من الأعمدة .

وسنوضح طريقة تحليل التشتت في هذه الحالة بمثال عددي .

جدول ١١ـ١٩ تجربة تصنيف ثنائي

متحول الأعمدة

		A	B	C	D	المجموع	المتوسط
متحول الصفوف	a	x_{11}	x_{12}	x_{13}	x_{14}	$T_{1.}$	$\bar{x}_{1.}$
	b	x_{21}	x_{22}	x_{23}	x_{24}	$T_{2.}$	$\bar{x}_{2.}$
	c	x_{31}	x_{32}	x_{33}	x_{44}	$T_{3.}$	$\bar{x}_{3.}$
المجموع		$T_{.1}$	$T_{.2}$	$T_{.3}$	$T_{.4}$	$T_{..}$	
المتوسط		$\bar{x}_{.1}$	$\bar{x}_{.2}$	$\bar{x}_{.3}$	$\bar{x}_{.4}$		$\bar{x}_{..}$

وفي الفقرات السابقة ، أي في حالة التصنيف الأحادي ، حصلنا على تقدير للتشتت من متوسطات الأعمدة . وسنقدر هنا التشتت من متوسطات الصفوف أيضاً ، وسنوضح الحسابات المطلوبة في تحليل التشتت من خلال البيان الإحصائي في الجدول (١١ـ ٢٠) . وبالنسبة لتقدير التشتت الذي نحصل عليه من متوسطات الأعمدة ، فإننا نحسبه كما في حالة التصنيف الأحادي تماماً ، أي أن مجموع مربعات الأعمدة (أو المعالجات) هو :

$$\frac{(13)^2}{3}+\frac{(16)^2}{3}+\frac{(17)^2}{3}+\frac{(14)^2}{3}-\frac{(60)^2}{12}=303.33-300=3.33.$$

جدول ١١ـ٢٠

	A	B	C	D	$T_{i.}$
a	7	6	8	7	28
b	2	4	4	4	14
c	4	6	5	3	18
$T_{.j}$	13	16	17	14	60

وبصورة مشابهة فإن مجموع المربعات المتعلق بمتوسطات الصفوف ، والذي يشكل بدوره تقديراً للتشتت هو :

$$\frac{(28)^2}{4} + \frac{(14)^2}{4} + \frac{(18)^2}{4} - \frac{(60)^2}{12} = 326 - 300 = 26.00$$

ومجموع المربعات الكلي هو :

$$7^2 + 2^2 + 4^2 + 6^2 + \ldots + 3^2 - \frac{(60)^2}{12} = 336 - 300 = 36.00$$

ونلاحظ أن هذه العلاقات كلها متشابهة ، فالمخرج يمثل دائماً عدد الملاحظات التي تشكل الصورة مجموعها . ويبين الجدول (١١ــ٢١) تحليل التشتت :

جدول ١١ــ٢١ تحليل التشتت

مصدر التغير	مجموع المربعات	درجات الحرية	متوسط المربعات
الصفوف	26.00	2	13.00
الأعمدة	3.33	3	1.11
الخطأ	6.67	6	1.11
المجموع	36.00	11	

ونحصل على مجموع مربعات الخطأ بطرح مجموع مربعات الصفوف ومجموع مربعات الأعمدة من مجموع المربعات الكلي . وبما أنه توجد ثلاثة صفوف فيكون لدينا درجتان من الحرية من أجل الصفوف . ويوجد أربعة أعمدة وبالتالي ثلاث درجات من الحرية من أجل الأعمدة . أما درجات الحرية الموافقة للخطأ فنحصل عليها بالطرح : 6=3−2−11 . ويمكن الحصول على النسبة F التي نختبر بواسطتها وجود فروق هامة بين الصفوف ، وذلك بقسمة متوسط مربعات الصفوف على متوسط مربعات الخطأ ، وهما تقديران

مستقلان للتشتت ، أي :

$$F = \frac{13.00}{1.11} = 11.7$$

وبالمقارنة مع $F_{.05}\ (2,6) = 5.14$ نرفض الفرضية بأنه لا توجد فروق بين متوسطات الصفوف . ويمكن القيام باختبار الفرق بين متوسطات الأعمدة ، وبصورة مستقلة عن وجود أو عدم وجود فروق بين الصفوف ، باستخدام النسبة :

$$F = \frac{1.11}{1.11} = 1.00,$$

والمقارنة مع $F_{.05}\ (3,6) = 4.76$ نقبل الفرضية بعدم وجود فروق بين متوسطات الأعمدة .

وقياساً على ما رأيناه في حالة التصنيف الأحادي نجد العلاقة :

$$\sum_{i=1}^{r}\sum_{j=1}^{c}(x_{ij} - \bar{x}_{..})^2 =$$

$$\sum_{i=1}^{r}\sum_{j=1}^{c}\left[x_{ij} - (\bar{x}_{i.} - \bar{x}_{..}) - (\bar{x}_{.j} - \bar{x}_{..}) - \bar{x}_{..}\right]^2$$

$$+ c\sum_{i=1}^{r}(\bar{x}_{i.} - \bar{x}_{..})^2 + r\sum_{j=1}^{c}(\bar{x}_{.j} - \bar{x}_{..})^2 \qquad (50)$$

ويمكن برهان هذه المعادلة بحساب كل من الحدود الأربعة في طرفيها . فكما وجدنا في الفقرة (١١ – ٣) يمكن كتابة الحد في الطرف الأيسر ، وهو مجموع المربعات الكلي ، على الشكل :

$$SST = \sum_{i=1}^{r}\sum_{j=1}^{c} x_{ij}^2 - \frac{T_{..}^2}{rc} \qquad (51)$$

ويمكن كتابة الحدين الأخرين من الطرف الأيمن على الشكل :

$$SSR = \frac{1}{c} \sum_{i=1}^{r} T_{i.}^{2} - \frac{T_{..}^{2}}{rc} \qquad (52)$$

$$SSC = \frac{1}{r} \sum_{j=1}^{c} T_{.j}^{2} - \frac{T_{..}^{2}}{rc} \qquad (53)$$

والحد الأول من الطرف الأيمن هو :

$$SSE = \sum_{i} \sum_{j} (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2$$

ويمكن كتابة هذا الحد على الشكل :

$$SSE = \sum_{i} \sum_{j} x_{ij}^{2} - \frac{\sum T_{i.}^{2}}{c} - \frac{\sum T_{.j}^{2}}{r} + \frac{T_{..}^{2}}{cr} \qquad (54)$$

ونرى بسهولة أن الأشكال الحسابية في (52) و (53) و (54) للحدود في الطرف الأيمن تجمع تماماً إلى الشكل الحسابي (51) للطرف الأيسر . وهكذا يمكن الحصول على مجموع المربعات SSE بطرح كل من مجموع مربعات الأعمدة SSC ومجموع مربعات الصفوف SSR من مجموع المربعات الكلي SST .

الخطوط العريضة لتحليل التشتت في حالة تصنيف ثنائي بملاحظة واحدة :

١ ـ الفرضية ١ : تأثيرات الأعمدة مساوية للصفر . ويتم اختبار هذه الفرضية مستقلاً عن تأثيرات الصفوف .

الفرضية ٢ : تأثيرات الصفوف مساوية للصفر . ويتم إختبار هذه الفرضية مستقلاً عن تأثيرات الأعمدة .

٢ ـ إختيار مستوى الأهمية α .

٣ ـ نستخدم الإحصاء F في إختبار الفرضيتين : من أجل الفرضية 1 نستخدم نسبة متوسط مربعات الأعمدة إلى متوسط مربعات الخطأ . ومن أجل الفرضية 2 نستخدم نسبة متوسط مربعات الصفوف إلى متوسط مربعات الخطأ .

٤ ـ وبفرض أن الملاحظات قد اختيرت بصورة عشوائية من مجتمعات طبيعية لها نفس التشتت وأن تأثيرات كل من الصفوف والأعمدة تجميعية ، يكون توزيع النسبتين F المذكورتين في الخطوة ٣ هما على الترتيب توزيع سنيديكور $F[c-1, (r-1)(c-1)]$ ، وتوزيع سنيديكور $F[r-1, (r-1)(c-1)]$

٥ ـ منطقة الرفض بالنسبة للفرضية 1 هي :

$$F > F_{1-\alpha}[c-1, (r-1)(c-1)]$$

ومنطقة الرفض بالنسبة لِلفرضية 2 هي :

$$F > F_{1-\alpha}[r-1, (r-1) c-1)]$$

٦ ـ نحسب النسبتين F ونرفض أو نقبل كلاً من الفرضيتين .

وعندما نصمم التجربة على هذا الشكل يمكن اختبار الفرضية بأن تأثيرات الأعمدة ولنرمز لها بِـ c_j مساوية للصفر وذلك مستقلاً عما إذا كان يوجد تأثيرات r_i للصفوف أم لا ؟ وبالمقابل يمكن إختبار الفرضية بأن تأثيرات الصفوف r_i مساوية للصفر وذلك مستقلا عما إذا كان يوجد تأثيرات c_j للأعمدة أم لا ؟ ويجب أن نفرض على أي حال أن هذه التأثيرات تجميعية ، أي أنه يمكن كتابة متوسط الخلية ij ، الخلية الموافقة للصف i والعمود j ، على الشكل $\mu_{ij} = \mu + r_i + c_j + e$ ، حيث μ نفسه بالنسبة لكل الخلايا أما r_i فيختلف من صف إلى آخر وهو بالتالي يمثل حصة أو عطاء الصف i في تعيين مقدار μ_{ij} ، وبصورة مشابهة فإن c_j التي تبقى نفسها بالنسبة لجميع خلايا العمود j تمثل مشاركة أو عطاء الصف i في تحديد

مقدار μ_{ij} ، أما ϵ فتدعى مركبة الخطأ . وإحدى تفسيرات خاصة التجميعية هو عدم وجود تفاعل بين الصفوف والأعمدة . أي أنه لا يوجد تأثير مشترك لصف وعمود يختلف عن تأثيريهما المنفصلين كما ذكرناهما أعلاه .

ويبين الجدول (١١–٢٢) جدول تحليل التشتت لتصنيف ثنائي بملاحظة واحدة في كل خلية ، حيث σ^2 هو التشتت المشترك لِـ rc من المجتمعات الموافقة للخلايا .

وتنبغي ملاحظة انه غالباً ما يجري تحليل التشتت دون التأكد التام من عدم وجود تفاعل بين متحولي الصفوف والأعمدة . ويمكن أن يقود مثل هذا التحليل إلى نتائج مفيدة إلا أنه ينبغي الحذر باعتبار أننا نعرف القليل عن تأثير التفاعل على نتائج التحليل .

١١–١٢ التصنيف الثنائي بعدة ملاحظات في الخلية الواحدة : نحصل على أفضل تقدير لتشتت الخطأ σ^2 من قياسات تتكرر تحت نفس الشروط . لنفرض أنه في دراسة إنتاج القمح لدينا تصنيف ثنائي وفقاً للمعالجات المطبقة (أسمدة مثلاً) ووفقاً للسلالات . وأننا زرعنا القمح في عدة وحدات تجريبية من أجل كل تركيب من التراكيب الممكنة بين المعالجات والسلالات . فنقول عندئذ أن التجربة قد أعيدت أو كررت . وسيمكننا مثل هذا التكرار من تحليل التجربة بصورة أكمل . والتحليل هو تركيب للطريقتين اللتين قدمناهما حتى الآن . فالبيان الإحصائي سيأخذ الشكل المبين في الجدول (١١–٢٣) مع فرض وجود صفين ، **ثلاثة** أعمدة ، **ثلاثة** تكرارات . (وسنستعرض هنا حالة عدد متساو من التكرارات في كل خلية ، ولا بد من تعديلات تطرأ على التحليل في حالة عدد غير متساو من التكرارات) .

جدول ١١-٢٢ تحليل التشتت لتصنيف ثنائي بملاحظة واحدة

مصدر التغير	مجموع المربعات	درجات الحرية	متوسط المربعات	النسبة F	توقع متوسط المربعات
ما بين الأعمدة	$\sum_{j=1}^{c} \frac{T_{.j}^2}{r} - \frac{T_{..}^2}{rc} = SSC$	$c-1$	$MSC = \frac{SSC}{c-1}$	$\frac{MSC}{MSE}$	$\sigma^2 + \frac{\sum_{j=1}^{c} c_j^2}{c-1}$
ما بين الصفوف	$\sum_{i=1}^{r} \frac{T_{i.}^2}{c} - \frac{T_{..}^2}{rc} = SSR$	$r-1$	$MSR = \frac{SSR}{r-1}$	$\frac{MSR}{MSE}$	$\sigma^2 + \frac{\sum_{i=1}^{r} r_i^2}{r-1}$
الخطأ	$SSE = SST - SSC - SSR$	$(c-1)(r-1)$	$MSE = \frac{SSE}{(c-1)(r-1)}$		σ^2
المجموع	$SST = \sum_{i=1}^{r}\sum_{j=1}^{c} x_{ij}^2 - \frac{T_{..}^2}{rc}$	$rc-1$			

جدول ١١ــ٢٦ تحليل التشتت

مصدر التغير	مجموع المربعات	درجات الحرية	متوسط المربعات	النسبة F
بين المتوسطات الستة	78.67	5	15.73	10.9
ما ضمن العينات	17.33	12	1.44	
المجموع	96.00	17		

وبمقارنة $F = 10.9$ مع $F_{.95}(5, 12) = 3.11$ نجد أنه توجد دلالة، عند المستوى .05، على أن متوسطات المجتمعات الستة غير متساوية. ونحلل الآن مجاميع الخلايا كما في حالة التصنيف الثنائي بملاحظة واحدة التي استعرضناها في الفقرة السابقة. ونثبت النتائج في الجدول (١١ــ٢٧).

ونحسب :

مجموع مربعات الصفوف :

$$SSR = \frac{38^2}{9} + \frac{(70)^2}{9} - \frac{(108)^2}{18} = 704.89 - 648 = 56.89$$

ونلاحظ أن كلاً من المجموعين 38 و 70 يحوي تسع ملاحظات. والمجموع الكلي 108 يحوي الملاحظات الثمانية عشر الأساسية المعطاة في الجدول (١١ــ٢٥).

مجموع مربعات الأعمدة :

$$SSC = \frac{41^2}{6} + \frac{(27)^2}{6} + \frac{(40)^2}{6} - \frac{(108)^2}{18} = 668.33 - 648 = 20.33$$

جدول ١١ــ٢٧ تحليل التشتت

مصدر التغير	مجموع المربعات	درجات الحرية	متوسط المربعات
ما بين الصفوف	56.89	1	56.89
ما بين الأعمدة	20.33	2	10.17
التفاعل بين الصفوف والأعمدة	1.45	2	0.72
ما بين المجاميع الجزئية	78.67	5	
ما ضمن العينات	17.33	12	1.44
المجموع	96.00	17	

مجموع المربعات من أجل « التفاعل » في الجدول (١١ــ٢٧) هو نفس مجموع مربعات الخطأ في تحليل التشتت الموافق للتصنيف الثنائي بملاحظة واحدة ، ونحصل عليه كما رأينا في الفقرة السابقة بالطرح أي :

$$78.67 - 56.89 - 20.33 = 1.45$$

وتُحدد درجات الحرية بنفس الطريقة أيضاً . أما مجموع مربعات المجاميع الجزئية فيوافق مجموع مربعات المتوسطات في الجدول (١١ــ٢٦) .

وقد استخدمنا حد التفاعل في هذا التحليل بدلاً من « الخطأ » باعتبار أنه يتوفر لنا تقدير آخر للتشتت من مجموع مربعات « ما ضمن العينات » يمكننا استخدامه لإختبار تلك الفروق بين المتوسطات التي لا يمكن أن نعزوها إلى فروق بين تأثيرات الصفوف أو فروق بين تأثيرات الأعمدة . ونعزو مثل هذه الفروق عادة إلى ما نسميه بالتفاعل بين الصفوف والأعمدة .

ونلاحظ أننا جزأنا مجموع المربعات الكلي إلى أربعة أجزاء : مجموع

جدول ١١ــ٢٨ تحليل التشتت لتصنيف ثنائي بعدة ملاحظات

مصدر التغير	مجموع المربعات	درجات الحرية	متوسط المربعات	F	توقع متوسط المربعات
ما بين الصفوف	$cn\sum_i(\bar{X}_{i..}-\bar{X}_{...})^2 = SSR$	$r-1$	$\frac{SSR}{r-1}=MSR$	$\frac{MSR}{MSE}$	$\sigma^2+cn\sigma_r^2$
ما بين الأعمدة	$rn\sum_j(\bar{X}_{.j.}-\bar{X}_{...})^2 = SSC$	$c-1$	$\frac{SSC}{c-1}=MSC$	$\frac{MSC}{MSE}$	$\sigma^2+rn\sigma_c^2$
التفاعل	$S_s-SSR-SSC = SSI$	$(r-1)(c-1)$	$\frac{SSI}{(r-1)(c-1)}=MSI$	$\frac{MSI}{MSE}$	$\sigma^2+n\sigma_I^2$
ما بين المجاميع الجزئية	$n\sum_i\sum_j(\bar{X}_{ij.}-\bar{X}_{...})^2 = S_s$	$rc-1$			
ما ضمن العينات (الخطأ)	$SST-S_s=SSE$	$rc(n-1)$	$\frac{SSE}{rc(n-1)}=MSE$		σ^2
المجموع	$\sum_i\sum_j\sum_k(X_{ijk}-\bar{X}_{...})^2 = SST$	$rcn-1$			

ونلاحظ من الجدول (١١ـ٢٨) أنه إذا كان $\sigma_I^2 = 0$ فإن متوسط مربعات التفاعل وما ضمن العينات يشكل كل منهما تقديراً منصفاً للتشتت σ^2 . وهذا يقترح علينا أن ضم التقديرين يمكن أن يقدم لنا تقديراً أفضل لِـ σ^2 ، وبالتالي يتيح فرصة لإختبار أدق بالنسبة لتأثيرات الصفوف والأعمدة . والقاعدة المقترحة عندئذ والتي لا ينقصها بعض التبرير النظري هو أن نقوم بعملية الضم هذه إذا كانت نسبة متوسط مربعات التفاعل إلى متوسط مربعات ما بين العينات أقل من ضعف قيمة $F_{.50}$ كما نحصل عليه من جدول التوزيع F . وباستخدام هذه القاعدة سنقوم بعملية الضم المقترحة باعتبار أن النسبة هي 0.72 وهي أقل من $2\ F_{.50}\ (2,12) = 2\ (.735) = 1.470$. وعملية الضم تتضمن جمع مجموع مربعات التفاعل إلى مجموع مربعات ما ضمن العينات للحصول على مجموع مربعات « الراسب » ، ونجمع أيضاً درجات الحرية للحصول على درجات الحرية الموافقة للراسب . وفي مثالنا أعلاه نجد أن متوسط الراسب هو $18.78/14 = 1.34$. ونستخدم الآن متوسط مربعات الراسب كمخرج لتشكيل النسب F الموافقة للصفوف والأعمدة . والنتائج مبينة في الجدول (١١ـ٢٩) .

جدول ١١ـ٢٩ تحليل التشتت

مصدر التغير	مجموع المربعات	درجات الحرية	متوسط المربعات	F
الصفوف	56.89	1	56.89	42.5
الأعمدة	20.23	2	10.17	7.59
الراسب	18.78	14	1.34	
المجموع	96.00	17		

وبما أن $F_{.05}\,(1,14) = 4.60$ فإننا نرفض الفرضية بتساوي متوسطات الصفوف . وباعتبار أن $F_{.05}\,(2,14) = 3.74$ فإننا نرفض أيضاً فرضية تساوي متوسطات الأعمدة وذلك عند المستوى $\alpha = .05$.

تمارين

١ ـ لدينا تحليل التشتت التالي :

مصدر التغير	درجات الحرية	مجموع المربعات	متوسط المربعات	توقع متوسط المربعات
بين المعالجات	4	244	61	$\sigma^2 + \frac{7}{4}\sum_{i=1}^{5} \tau_i^2$
بين الوحدات التجريبية ضمن المعالجات	30	270	9	σ^2
المجموع	34	514		

أ ـ اكتب النموذج المناسب .

ب ـ اعرض الفرضية الابتدائية التي صُممت التجربة لإختبارها .

حـ ـ اختبر الفرضية المعطاة في (ب) عند مستوى الأهمية $\alpha = .05$.

٢ ـ ليكن التحليل التالي الناتج عن تجربة تحوي 6 معالجات ، 10 وحدات تجريبية من أجل كل معالجة ، وثلاثة قياسات من أجل كل وحدة تجريبية :

مصدر التغير	درجات الحرية	متوسط المربعات	توقع متوسط المربعات
المعالجات	5	12489	$\sigma_\delta^2 + 3\sigma^2 + \frac{30}{5}\sum_{i=1}^{6}\tau_i^2$
الوحدات التجريبية ضمن المعالجات	54	3339	$\sigma_\delta^2 + 3\sigma^2$
القياسات ضمن الوحدة التجريبية	120	627	σ_δ^2
المجموع	179		

أ ـ اكتب النموذج الموافق للتجربة عارضاً بوضوح ما يمثله كل حد من الحدود .

ب ـ اختبر الفرضية بأن للمعالجات الست نفس المتوسط .

حـ ـ احسب تشتت متوسط المعالجة .

د ـ اذا علمنا أن متوسط المعالجة الثالثة هو 193.7 فاحسب وفسرّ % 95 مجال ثقة للمتوسط الحقيقي للمجتمع الموافق للمعالجة الثالثة .

٣ ـ هدف تجربة هو حساب مركبة التشتت الموافقة لتغيرات تركيز حامض الاسكوربيك (ملغ في المائة غرام) في أوراق اللفت . وقد أخذت ورقتان من المنطقة المركزية من كل من خمس نبتات وقيس تركيز حامض الاسكوربيك في كل ورقة . وقد كُرّرت هذه العملية يومياً على مدى ستة أيام ، علماً أن اختياراً جديداً من النبتات يتم في كل يوم . وكانت النتائج كما نجد في البيان الإحصائي التالي :

اليوم	الورقة	النبتة				
		1	2	3	4	5
1	A	9.1	7.3	7.3	10.7	7.7
	B	7.3	9.0	8.9	12.7	9.4
2	A	12.6	9.1	10.9	8.0	8.9
	B	14.5	10.8	12.8	9.8	10.7
3	A	7.3	6.6	5.2	5.3	6.7
	B	9.0	8.4	6.9	6.8	8.3
4	A	6.0	8.0	6.8	9.1	8.4
	B	7.4	9.7	8.6	11.2	10.3
5	A	10.8	9.3	7.3	9.3	10.4
	B	12.5	11.0	8.9	11.2	12.0
6	A	10.6	10.9	10.4	13.1	7.7
	B	12.3	12.8	12.1	14.6	9.4

٤ ـ اذا علمنا أن متوسطات 10 أفراد في كل من خمس مجموعات

هي 30, 34, 34, 36, 38 و أن تشتت متوسط المجموعة هو 8 ، فاكتب جدول تحليل التشتت .

٥ ـ ليكن تحليل التشتت التالي :

مصدر التغير	درجات الحرية	مجموع المربعات	متوسط المربعات	توقع متوسط المربعات
المعالجات	3	1800	600	$\sigma_\delta^2 + 3\sigma^2 + 30\sigma_\tau^2$
عينات ضمن المعالجات	36	4320	120	$\sigma_\delta^2 + 3\sigma^2$
قياسات ضمن العينات	80	960	12	σ_δ^2
المجموع	119	7080		

أ ـ احسب تشتت متوسط المعالجة .

ب ـ اختبر الفرضية $H_0 : \sigma_\tau^2 = 0$ وفسّر جوابك .

جـ ـ إذا علمنا أن متوسط العينة من المعالجة 1 هو 80 ، فاحسب % 95 مجال ثقة لتقدير المتوسط الحقيقي للمعالجة رقم 1 .

٦ ـ كل مجموعة من مجموعات الملاحظات التالية هي عينة عشوائية من مجتمع طبيعي . اختبر تجانس التشتتات . ثم اختبر تساوي المتوسطات باستخدام تحليل التشتت

A	B	C	D
49	49	44	58
42	44	57	54
47	50	34	64
76	58	48	60
69	70	50	53
58			64
			52
			42

٧ ـ ترغب وكالة في تحديد ما إذا كانت خمسة أنواع من السيارات تقطع نفس عدد الأميال من أجل كل غالون من الوقود . ومن أجل ذلك استخدمت ثلاث سيارات من كل نوع وذلك من كل من ثلاث مدن ، واختبرت ما تقطعه كل عربة من أجل غالون من البنزين . وكانت النتائج كما في الجدول .

أ ـ لماذا اخترنا ثلاث مدن وليس مدينة واحدة ؟

ب ـ ما هي المجتمعات التي جاءت منها العينات ؟

جـ ـ ماذا نفترض بالنسبة للمجتمعات وما هي الفرضيات التي يمكن اختبارها ؟

د ـ أنجز تحليل التشتت واعرض نتائجك بالكامل .

المدن

	لوس انجلوس			سان فرانسيسكو			بورتلاند		
A	20.3	19.8	21.4	21.6	22.4	21.3	19.8	18.6	21.0
B	19.5	18.6	18.9	20.1	18.9	20.5	19.6	18.3	19.8
C	22.1	23.0	22.4	20.1	21.0	19.8	22.3	22.0	21.0
D	17.6	18.3	18.2	19.5	19.2	20.3	19.4	18.5	19.0
E	23.6	24.5	25.1	17.6	18.3	18.1	22.1	24.3	23.8