٩ تحليل التباين أحادى الاتجاه

One way analysis of variance

١- مقدمة

٢- المتغير والعامل والمعاملة والمستوى

٣- افتراضات نموذج تحليل التباين

٤- ماذا يراد عادة من تحليل التباين ؟

٥- مفهوم وكيفية إجراء تحليل التباين

٦- النموذج الرياضى

٧- اختبار جميع الفروق بين أزواج المتوسطات

٨- المقارنات المستقلة (المتعامدة)

٩- تحليل التباين عند عدم تساوى المكررات

١٠- النموذج التجميعى الخطى

١١- تقدير مكونات التباين فى حالة اختلاف حجم العينات

١٢- معامل الارتباط الداخلى (الجوانى)

١٣- العينات داخل العينات

٩-١ مقدمة

تنحصر أهم الأسس فى دراسة الموضوعات ذات الطابع العلمى فى دراسة الاختلافات بين الظواهر المختلفة. فبدون الاختلافات أو التباينات لا توجد طريقة لدارسة العلاقة بين المتغيرات، أى أن دراسة التباين variance فى البيانات يعطى دلالة على مدى تشتتها وفى نفس الوقت أيضاً يوضح مدى اتساق هذه البيانات أو تجانسها homogenous من عدم اتساقها hetcrogeneous.

٩-٢ المتغير والعامل والمعاملة والمستوى
Variable, factor, treatment, level

فى الواقع المتغير والعامل والمعاملة والمستوى هى ألفاظ ذات معانى متداخلة، فالمتغير هو المشاهدة التى قد تأخذ قيماً مختلفة. وقد يكون هذا المتغير مستقلاًindependcnt variable، أى لا يفترض أنه يتأثر بمتغير آخر فى نطاق التجربة. أو قد يكون المتغير تابعاً dependent variable، أى يتأثر بمتغير آخر فى نطاق التجربة، وقد يطلق عليه أحياناً متغير الاستجابة response variable.

أما العامل فهو كثيراً ما يستخدم بمفهومه العام مثل قول العوامل المؤثرة على الإنتاج، العوامل المؤثرة على خصوبة الحيوان ... الخ. ولكن فى المجال الإحصائى فالعامل يقصد به مؤثر (أى متغير مستقل) له عدة مستويات levels، مثال ذلك إذا وضع جنس الحيوان فى نموذج model حيث المتغير التابع هو النمو والمتغير المستقل هو الجنس، والجنس فى هذه الحالة يعتبر عاملاً يأخذ عدة مستويات: ذكر، أنثى، مخصى ...الخ. وقد يكون العامل متغيراً يرغب المجرب فى دراسة أثره على متغير تابع مثل دراسة تأثير مستويات مختلفة من البروتين فى العليقة على نمو الحيوان. وقد يرغب المجرب فى دراسة أثر أكثر من عامل فمثلاً أثر البروتين (P) وله ثلاثة مستويات وأثر إضافة فيتامين معين (V) وله مستويين. فى هذه الحالة هناك عاملان هما البروتين والفيتامين وتجربة كل من مستويات البروتين مع كل من مستويات الفيتامين يطلق عليه تجارب عاملية factorial experiments. والتجربة فى الواقع عبارة عن 6 معاملات هى $P_1V_1, P_1V_2, P_2V_1, P_2V_2, P_3V_1, P_3V_2$ أو معاملة واحدة لها 6 مستويات. إذن المعاملة هى عامل له مستويات ولكن ليس كل عامل معاملة حيث ممكن أن يكون العامل هو الموقع أو العمر أو الحظيرة، وإن كانت هذه كلها يمكن جعلها عاملاً إذا قصدت مباشرة بالدراسة ولها مستوياتها المقصودة بالبحث. إذا يمكن القول أن كل عامل أو معاملة متغير ولكن ليس كل متغير عاملاً أو معاملة. وسوف تستخدم هذه التعبيرات تبادليا فى هذا المؤلف.

هذا وقد سبقت الإشارة فى الباب الثامن إلى استخدام التباين أو مشتقاته فى اختبار الفرض الخاص بتساوى متوسطى عشيرتين. حيث تم استخدام اختبار فرض العدم بواسطة اختبار(t). ولكن فى كثير من الحالات يفترض وجود أكثر من عشيرتين ويراد اختبار الفرض الخاص بالفروق فيما بينها. فمثلا قد يكون الاهتمام مركزا فى اختبار اختلاف مستويات مكون ما من العليقة على نمو الحيوانات. فإذا افترض وجود أكثر من عشيرتين، فإنه يمكن إجراء سلسلة من اختبارات (t) فى كل منها يتم اختبار الفرق بين متوسط عشيرة ما والمتوسطات الأخرى. فعلى سبيل المثال لو كانت هناك ثلاثة متوسطات (ثلاثة عشائر مفترضة) فسوف يتم عمل ثلاث اختبارات، ولو كان هناك 4 متوسطات يتم عمل ستة اختبارات للفروق، ... وهكذا تتزايد هذه الاختبارات بزيادة عدد المتوسطات المختبرة. وفى نفس الوقت فإن كل اختبار (t) سوف يكون مرتبطاً بمستوى قدره α لاحتمال حدوث خطأ من النوع الأول type I error، وعلى هذا الأساس فإنه إذا أجرى العديد من هذه الاختبارات فإنه من المتوقع الوصول إلى قرار خاطئ فى بعض من هذه الاختبارات متوقفاً على درجات الاحتمال.

وتحليل البيانات بطريقة تحليل التباين يعتبر امتداداً لاختبار(t) والخاص باختبار فرض العدم عن المتوسطات. وهذه الطريقة تسمح باختبار واحد وباحتمال خطأ واحد حجمه α أن يجرى اختبارات تجيب على الأسئلة الخاصة بالمتوسطات واختلافها، كأن يكون السؤال عما إذا كانت البيانات توضح أن العشائر المفترضة تختلف فيما بينها، وأيضاً إذا كانت هذه الاختلافات تعتبر معنوية ولا ترجع إلى الصدفة.

ويتلخص الغرض من تحليل التباين فى تقسيم التباين الكلى بين البيانات إلى مكونات ذات مغزى والتى تقيس مصادر الاختلافات.

ففى حالة اختبار الفروق بين المعاملات الموجودة فى العليقة وتأثيرها على اختلاف النمو فى الأفراد، فإن تحليل التباين سوف يقوم بتقسيم التباين الكلى إلى جزء يرجع إلى الخطأ التجريبى experimental error والجزء الثانى يرجع إلى الخطأ التجريبى بالإضافة إلى أي تباين يرجع إلى اختلاف مستويات العامل فى العليقة. فإذا كان فرض العدم صحيحاً، أى أنه توجد اختلافات بين معدلات النمو على المستويات المختلفة للعليقة، فإن كلا من الجزئين المقسم إليهما التباين يعطى تقديراً مستقلا للخطأ التجريبى. وبالتالى فإن الاختبار يقوم على أساس مقارنة لكل من مصدرى التباين أو الاختلاف، وأن مستوى معنوية هذا الاختلاف سوف يوضح عن طريق نسبة التقديرين للتباين لبعضهما، وهذه تتبع توزيعاً يعرف بتوزيع ف F distribution.

الأساس الرياضى لتحليل التباين قدمه أحد أئمة علم الإحصاء وهو العالم الإنجليزى R. A. Fisher، ولقد قام العالم الأمريكى G. W. Snedecor بدراسة

الناحية الرياضية للنسبة بين تقديرات التباين وتوزيعها وسمى بتوزيع F نسبة إلى اسم Fisher.

٩-٣ افتراضات نموذج تحليل التباين Assumption in analysis of variance

إذا أريد دراسة أثر الزيادة فى كمية البروتين فى العليقة والتى ستكون مصحوبة بنمو أسرع، فإن فرض العدم هنا هو أن متوسط النمو فى الحيوانات التى لم تتلق المادة فى المعاملة تتساوى مع تلك التى حصلت على المادة، وبدرجات متزايدة، أى أن جميع المتوسطات للمجموعات التى تلقت المعاملات المختلفة متساوية.

يعتبر تحليل التباين أساساً اختباراً للمعنوية، حيث يكون المطلوب، بناء على العينة، هو تقدير تواجد من عدم تواجد علاقة بين المعاملة والاستجابة فى العشائر المختلفة. وهناك بعض الافتراضات فى عملية أخذ العينة تعتبر أساسية حتى يصبح تحليل التباين صحيحاً من ناحية الاستقراءات التى سيتم الوصول إليها من العينة. وهذه الافتراضات هى:

أولا: أن الأفراد فى العينة يجب أن تكون مستقلة independent وعشوائية random فى اختيارها. وهذا يعنى أن هناك استقلالية فى كون أى فرد يدخل ضمن العينة وهذا أيضاً يوضح أنه ليست هناك علاقة بين اختيار فرد ما فى العينة واختيار فرد آخر. ومن خصائص العينات العشوائية ليس فقط أن الأفراد لها نفس الفرصة فى الظهور فى العينة ولكن أيضاً إعطاء أى توليفة من الأفراد نفس الفرصة للاختيار.

ثانياً: تجانس التوزيع أو الانتشار وهو ما يعرف homoscedasticity بمعنى آخر أن العشائر المختلفة المأخوذ منها العينة لها نفس التباين common variance. وهذا يعنى أنه، حتى ولو وجدت اختلافات بين المتوسطات فى العشيرة المؤثر عليه بالمعاملات فإن التشتت أو التباين داخلها لم يتغير وأن تباينات القيم غير مرتبطة مع بعضها.

ثالثاً: حيث إن تحليل التباين يفترض إجراء اختبارات معنوية وحتى يمكن إجراء مثل هذه الاختبارات فإنه يفترض أن الأفراد والمشاهدات المأخوذة من عشيرة واحدة تختلف عن بعضها اختلافاً طبيعياً (الأخطاء) وأن هذه الاختلافات مأخوذة من عشيرة تتبع التوزيع الطبيعى normal distribution بمتوسط صفر وتباين قدره σ^2.

وعموماً فإن هذه الافتراضات إذا لم تتحقق violated بوضوح فإن تحليل التباين وبالتالى اختبار المعنوية سوف يؤدى إلى الوصول إلى نتائج أو قرارات متحيزة

ويصبح هناك عدم ثقة فى الاستقراءات المأخوذة لتطبيقها على العشيرة حيث يمثل الهدف الأساسى من استخدام العينة وتحليل نتائجها. وسوف تتم مناقشة أهم هذه الافتراضات وطرق معالجتها فى حالة عدم تحققها فى الباب الثانى عشر.

٩-٤ ماذا يراد عادة من تحليل التباين؟

عند أخذ الحالة السابقة والخاصة بدراسة تأثير إضافة مادة ما إلى العليقة على نمو الأفراد، فإذا كان هناك ثلاثة مستويات لهذه المادة وتم توزيع أفراد العينة عشوائيا على المستويات الثلاثة فإن ذلك سوف يؤدى إلى وجود ثلاثة متوسطات يمثل كل منها عشيرة تتلقى نوعاً من المعاملة، والمطلوب معرفة ما إذا كانت هذه العشائر تختلف عن بعضها علماً بإنه من المفترض أن تأثير المعاملة سوف لا يؤثر على التباينات بين مجموعة الأفراد التى تتلقى نفس المعاملة. وبالتالى يوجد ثلاثة متوسطات يرمز لها بالرمز $\overline{Y}_i$ حيث i تمثل مستوى المعاملة $i = 1,2,3$.

وهناك ثلاث حالات يمكن تصويرها بيانياً فى الشكل ٩-١ (أ، ب، ج) والذى يبين العلاقة الممكن تصورها بين مجموعة من العينات أو المواقف والعشيرة المأخوذ منها هذه العينات.

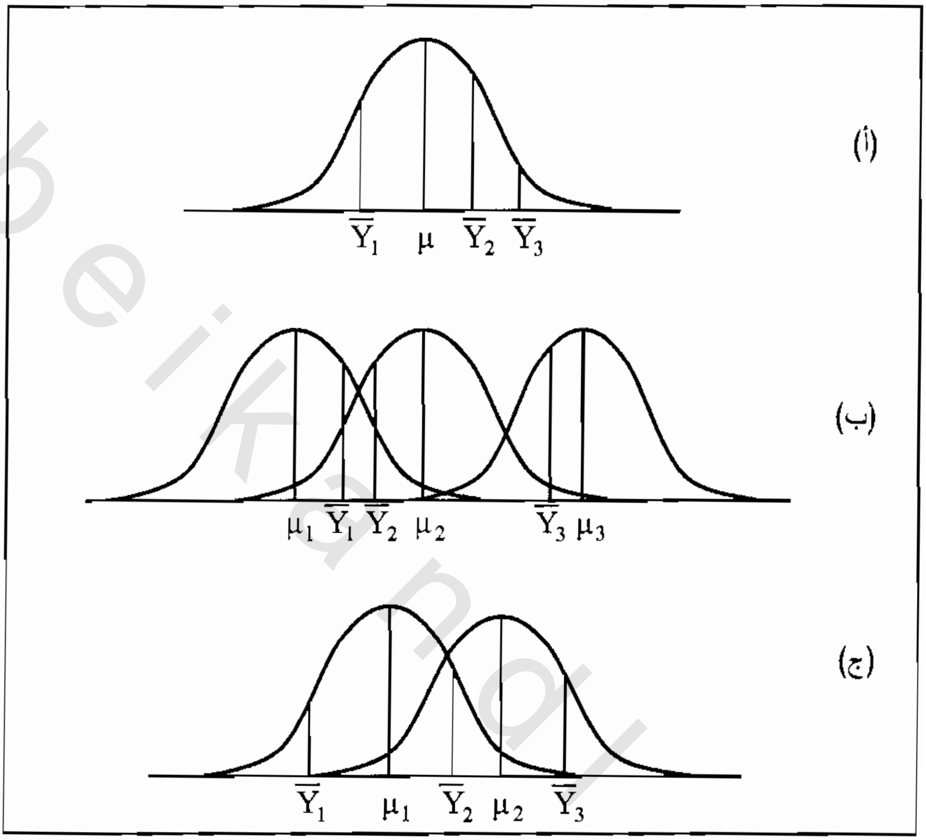
الحالة (أ): يوجد توزيع طبيعى واحد وهذا يمثل عشيرة واحدة. أما بالنسبة للقيم المختلفة الموضحة للمتوسط $\overline{Y}_i$ فإنها تمثل متوسطات العينات أو المعاملات الثلاثة والتى تعتبر مختلفة فقط للأسباب الطبيعية لاختلاف نقطة التقدير point estimate والتى تحسب من كل عينة لتمثيل متوسط العشيرة. أى أنه ليس هناك سبب للقول بأن هذه النقاط تعتبر مختلفة عن بعضها ولكنها جميعاً مأخوذة من نفس العشيرة والتى متوسطها μ.

الحالة (ب): فهى مختلفة حيث إن هناك ثلاث عشائر كل منها ممثل بتوزيع طبيعى وقيم $\overline{Y}_3$، $\overline{Y}_2$، $\overline{Y}_1$ تعتبر نقاطاً تقديرية لمتوسط الثلاثة عشائر المختلفة μ_3، μ_2، μ_1. وبما أن الثلاثة متوسطات للعينات (المعاملات) مختلفة عن بعضها فإن هذا يعنى وجود اختلافات بين متوسطات العشائر الثلاثة.

الحالة (ج): هناك متوسطان لعينتين تمثلان متوسطاً لعشيرة واحدة μ_1 فى حين أن متوسط المعاملة الثالثة $\overline{Y}_3$ فإنه يعتبر تقديراً لمتوسط العشيرة μ_2 وهو يختلف عن الأول كنتيجة لتأثير المعاملة.

ومن نفس المنطلق فإنه يمكن توسيع مدى فرض العدم ليشمل أى عدد من المتوسطات المفترضة أو بالتالى أى عدد من مستويات العامل. وفى كل هذه الحالات فإن الفرض المقابل هو أنه ليست كل المتوسطات متساوية. وعلى هذا الأساس يتم

رفض فرض العدم فى حالة اختلاف ولو واحد فقط من متوسطات العشيرة عن بقية القيم. وهذا هو أساس تحليل التباين وبالتالى اختبار المعنوية.



شكل ٩-١ العلاقة الممكن تواجدها بين مجموعة من العينات والعشيرة المأخوذ منها العينات

ويمكن تحليل التباين من استخدام كل البيانات الموجودة فى التجربة (الدراسة) دون النظر إلى المسبب، لتقسيم التباين الكلى إلى تأثير العامل المختبر (المعاملات) والخطأ التجريبى. ثم بعد ذلك مقارنة مصادر التباين بواسطة اختبار F.

٩-٥ مفهوم وكيفية إجراء تحليل التباين

سوف يستخدم المثال العددى التالى لتوضيح مفهوم وكيفية إجراء تحليل التباين والبيانات تمثل 4 معاملات مختلفة لحفظ مادة معينة (قد تكون المعاملات الأربعة

عبارة عن 4 مواد حفظ مختلفة أو 4 مستويات مختلفة من نفس مادة الحفظ) وتمثل الأرقام عدد الساعات التى تبقى فيها المادة صالحة قبل أن تفسد:

المكررات	مادة الحفظ أ	ب	ج	د
١	7	5	3	10
٢	8	3	4	7
٣	7	3	3	9
المجموع	**22**	**11**	**10**	**26**
المتوسط	**7.33**	**3.67**	**3.33**	**8.67**

فإذا افترض أنه لا توجد اختلافات بين الأربعة معاملات فإن 12 مشاهدة تتوزع حول متوسط واحد μ وتباين مشترك فيما بينها σ^2.

وكما سبق فإنه يمكن من خلال تحليل التباين الحصول على أكثر من تقدير للتباين فى العشيرة. وبالتالى يوجد ثلاثة تقديرات يمكن الحصول عليها لتقدير التباين σ^2.

أولاً: إذا افترض أن فرض العدم صحيح، فإن هذا يعنى أن كل البيانات الموجودة فى التجربة مأخوذة من نفس العشيرة. وبالتالى يمكن الحصول على تقدير للتباين الكلى من مجموع المربعات الكلى المصحح وهى:

$$\sum y^2 = \sum Y_i^2 - \frac{(\sum Y)^2}{n} = (7^2 + 8^2 + \cdots + 9^2) - \frac{(69)^2}{12}$$
$$= 469 - 396.75 = 72.25$$

ومجموع المربعات المصحح هذا له 11 درجة حرية. وبالتالى يصبح متوسط المربعات $72.25/11 = 6.568$ وهو أول تقدير للتباين σ^2.

ثانياً: التقدير الثانى للتباين فى العشيرة σ^2 يمكن الحصول عليه عن طريق حساب مجموع المربعات المصحح لكل من الأربعة معاملات بنفس الطريقة التى سبق شرحها فى الباب الثامن وحساب ما سبق وأطلق علية التباين المجمع pooled variance، وبالتالى:

$$S_p^2 = \sum y_1^2 + \sum y_2^2 + \sum y_3^2 + \sum y_4^2 \qquad (٩-١)$$
$$= 0.67 + 2.67 + 0.67 + 4.67 = 8.68$$

عدد درجات حرية هذا المجموع هو $t(n-1)$ حيث t هى عدد المعاملات 4، n عدد الأفراد (المكررات داخل المعاملة)، أى أن درجات الحرية 8 درجات حرية. وقسمة مجموع المربعات هذا على درجات الحرية يعطى تقدير $\sigma^2 = 1.085$ من داخل المعاملات. ويعرف هذا بالتباين داخل المعاملات within treatment mean square.

وطبعاً من الممكن ملاحظة أنه فى كل حالة تم حساب مجموع انحرافات أفراد العينة عن متوسطها. أى أنه فى هذه الحالة لا يتأثر هذا التقدير باختلاف متوسطات المعاملات أو عدم اختلافها. وبما أن هذه القيم مقاسة على الأفراد، أى كل وحدة تلقت المعاملة ككل فإن تقدير التباين من داخل المعاملات يعتبر قياساً للخطأ التجريبى للتجربة.

ثالثاً: أما التقدير الثالث والأخير فهو الذى يقيس الاختلافات بين المعاملات أو تأثيراتها ككل. إن متوسطات المعاملات يعتبر كل منها تقديرا لمتوسط العشيرة. وهذه المتوسطات لها تباين حول متوسط العشيرة يساوى σ^2/n حيث n هى عدد المشاهدات داخل كل معاملة (عينة). وبالتالى فإنه إذا تم حساب مجموع المربعات للانحرافات المأخوذة من متوسطات العينات وبضربه فى n يعطى مجموع مربعات الانحرافات بين هذه المتوسطات والمتوسط العام، وفى المثال فإن متوسطات المعاملات هى 7.33, 6.67, 3.33,8.67 وبالتالى فإن مجموع مربع الانحرافات يمكن حسابه كالتالى:

$$(3)\left[(7.33)^2 + (3.67)^2 + (3.33)^2 + (8.67)^2 - \frac{(23)^2}{4}\right]$$
$$= (3)(21.2056) = 63.6168$$

ومجموع المربعات المحسوب من المتوسطات له 3 درجات حرية وبالتالى فإن تقدير التباين من المتوسطات يكون $63.6168/3 = 21.21$.

وعادة ما يتم حساب مجموع المربعات بين المعاملات من معادلة (٩-٤) والتى سيأتى ذكرها فيما بعد، وعلى ذلك تصبح العمليات الحسابية لحساب مجموع المربعات المصحح بين المعاملات كالتالى:

$$\frac{(22)^2 + (11)^2 + (10)^2 + (26)^2}{3} - \frac{(69)^2}{12}$$
$$= \frac{1381}{3} - \frac{4761}{12} = 63.583$$

وهى تساوى نفس القيمة المذكورة سابقاً باستثناء بعض أخطاء التقريب. وعموما فإن المعادلة (٩-٤) هى الأكثر دقة وذلك لعدم إجراء التقريب لكل متوسط عند حساب المربعات منها. كما أنه عادة، يتم حساب مجموع المربعات داخل المعاملات بطرح المربعات بين المعاملات من مجموع المربعات الكلى فى هذا النوع من تحليل التباين وذلك نتيجة لخاصية مهمة فى تحليل التباين وهى خاصية التجميع لمجاميع المربعات concept of additivity of sums of squares and وأيضا لدرجات الحرية المقابلة degrees of freedom (هذه الخاصية تظل ثابتة وصحيحة فى تحليل البيانات للتقسيم الأحادى وأيضا فى حالة التقسيمات الأخرى المتعددة، إذا كانت الأعداد فى الفئات إما متساوية أو متناسبة proportional، ولكن هذه الخاصية تفقد صلاحيتها إذا اختلفت التكرارات فى التقسيمات المتعددة مما يستوجب إجراء تعديلات على تحليل البيانات الخاصة بتحليل التباين).

وعلى هذا الأساس فإن مجموع المربعات داخل المعاملات هو مجموع المربعات الكلى مطروحاً منه مجموع المربعات بين المعاملات.

نظرية ٩-١

$$\sum_i \sum_j (Y_{ij} - \overline{Y}_{..})^2 = n\sum_i (\overline{Y}_{i.} - \overline{Y}_{..})^2 + \sum_i \sum_j (Y_{ij} - \overline{Y}_{i.})^2$$

إثبات النظرية

بإضافة $\overline{Y}_{i.}$ مرة وطرحه مرة أخرى وإعادة الترتيب ينتج أن:

$$\sum_i \sum_j (Y_{ij} - \overline{Y}_{..})^2 = \sum_i \sum_j \left[(\overline{Y}_{i.} - \overline{Y}_{..}) + (Y_{ij} - \overline{Y}_{i.})\right]^2$$

$$= \sum_i \sum_j \left[(\overline{Y}_{i.} - \overline{Y}_{..})^2 + 2(\overline{Y}_{i.} - \overline{Y}_{..})(Y_{ij} - \overline{Y}_{i.}) + (Y_{ij} - \overline{Y}_{i.})^2\right]$$

$$= \sum_i \sum_j (\overline{Y}_{i.} - \overline{Y}_{..})^2 + 2\sum_i \sum_j (\overline{Y}_{i.} - \overline{Y}_{..})(Y_{ij} - \overline{Y}_{i.}) + \sum_i \sum_j (Y_{ij} - \overline{Y}_{i.})^2$$

وبما أن مجموع الانحرافات على كل المدى يساوى صفر، فإن الحد الأوسط يصبح صفر حيث إن:

$$\sum_i (\overline{Y}_{i.} - \overline{Y}_{..}) = \sum_j (\overline{Y}_{ij} - \overline{Y}_{i.}) = 0$$

لاحظ أن الحد الأول لا يحتوى على أفراد العينة ولكنه عبارة عن مربع الفرق بين متوسط كل معاملة والمتوسط العام، وبالتالى يمكن كتابته على النحو التالى:

$$\sum_i \sum_j (\overline{Y}_{i.} - \overline{Y}_{..})^2 = n \sum_i (\overline{Y}_{i.} - \overline{Y}_{..})^2$$

وبالتالى فإن:

$$\sum_i \sum_j (Y_{ij} - \overline{Y}_{i.})^2 = \sum_i \sum_j (Y_{ij} - \overline{Y}_{..})^2 - n \sum_i (\overline{Y}_{i.} - \overline{Y}_{..})^2 \qquad (٩-٢)$$

حيث i هى رتبة المعاملات وعددها t فى حين أن j هو رتبة الأفراد أو المشاهدات داخل كل معاملة وعددها n. ففى النظرية (٩-١) نجد أن الحد الأيسر هو مجموع المربعات الكلية total sum of squares أما الجانب الأيمن فحده الأول هو مجموع المربعات بين المعاملات (الفئات) between classes sum of squares، والحد الثانى هو مجموع المربعات داخل المعاملات within classes sum of squares، وهذا الأخير يطلق عليه فى كثير من الحالات الخطأ error ويعنى الخطأ التجريبى experimental error.

٩-٦ النموذج الرياضى Mathematical model

لقد سبق الحديث عن الافتراضات التى يجب توافرها ليصبح تحليل التباين صحيحاً. هذه الافتراضات تختص بالمشاهدات التى يجرى تحليلها. ويمكن تعريف (وصف) كل مشاهدة بنموذج رياضى يحوى تلك التأثيرات التى يفترض أنها تؤدى إلى حدوث الاختلافات بين المشاهدات بعضها مع البعض. والنموذج الرياضى للتقسيم أحادى الجهة يعبر عنه كالتالى:

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

حيث:

μ هى المتوسط العام للعشيرة المأخوذ منها هذه المشاهدات، كأن يكون عشيرة أوزان الميلاد للعجول من سلالة معينة أو نسبة البيض الذى يفقس من سلالة معينة من الدجاج أو إجمالى الدخل أو الإنفاق الأسرى شهرياً أو سنوياً فى محافظة معينة ... الخ. أى أن μ تمثل الجزء المشترك بين كل الأفراد التى ينطبق عليها تعريف هذه العشيرة.

α_i تعبر عن تأثير الفئة (المستوى) class أو المعاملة والمفترض أنها تؤثر على كل الأفراد التى تتبع أو الواقعة تحت تأثير هذا العامل و i تمثل عدد تلك الفئات ويأخذ القيم المختلفة من 1 إلى t العدد الكلى لتلك الفئات. وعلى هذا الأساس فإنه يفترض أن متوسط أى معاملة أو فئة $\mu_i = \mu + \alpha_i$ وأن القيمة المتوسطة لجميع هذه الفئات أى متوسط قيم μ_i يساوى μ، المتوسط العام. ويتبع ذلك أن مجموع قيم α_i يساوى الصفر أى أن $\sum \alpha_i = 0$، أى أن كل قيمة من قيم α_i هى عبارة عن انحراف متوسط الأفراد التابعة لهذه الفئة عن المتوسط العام.

$\in_{ij}$ هذا الجزء من النموذج تم تسميته من قبل بالخطأ والذى افترض أنه مأخوذ من توزيع طبيعى للأخطاء (الاختلافات) بمتوسط قدره صفر وتباين σ^2 ويرمز لذلك بأن $\in_{ij} \sim NID(0, \sigma^2)$ أى أنها تتوزع مستقلة عن بعضها وطبيعيا حول متوسط للاختلافات قدره صفر ونفس التباين σ^2.

Y_{ij} تعبر عن المشاهدة التى تحت فئة (مستوى) المعاملة i وترتيبها j بين المكررات والتى عددها n.

ومن الطبيعى أن الفئات المختلفة لها متوسطات مختلفة μ_i ولكن الاختلافات أو التباينات داخل كل فئة يفترض أن لها قيمة واحدة هى σ^2.

وهذه الطريقة فى وصف المشاهدة Y_{ij} تقترح أيضاً أنه إذا لم يكن هناك تأثير للمعاملات، أى أن μ_i كلها متساوية، فإن النموذج الرياضى يختزل إلى الصورة الأكثر سهولة، أى إذا كان فرض العدم صحيحاً وهى:

$$Y_i = \mu + \in_i \quad , i = 1, 2, \ldots, tn$$

وبالتالى فإن تقسيم الأفراد أو المشاهدات إلى أقسام أو فئات كان تقسيماً عشوائياً لا يدل فى الواقع على وجود اختلافات بين هذه الفئات ويصبح المصدر الوحيد للتباينات هى تلك الاختلافات الطبيعة بين الأفراد داخل العشيرة ككل.

كان الاهتمام فى هذا النموذج ينحصر فى تأثير المعاملات، فهل هذه المعاملات تؤدى إلى وجود اختلاف فى المتوسطات أم لا؟ أى هل ينشأ عن ذلك عشائر مختلفة كل منها له متوسط مختلف μ_i؟ ولاستبيان ذلك فإن الاهتمام يكون بالأساس إجراء اختبار للمعنوية لتوضيح ما إذا كانت كل المتوسطات متساوية أم لا، وهو ما سبق إيضاحه أنه يتم بواسطة اختبار F.

لاستخلاص النتائج drawing inference من النموذج الرياضى أو بمعنى آخر لاستقراء النتائج يجب أن يكون واضحاً لدى المجرب كيفية اختيار المعاملات أو الأقسام المعبر عن أثرها بـ α فى النموذج لأن كيفية الاختيار هذه ستحدد الاستنتاج المنشود، وللوصول إلى ذلك يلزم التفرقة بين الأثر الثابت fixed effect والأثر العشوائى random effect.

صندوق ٩-١

- تحليل التباين هو وسيلة لتقسيم التباين الكلى المشاهد وإرجاع كل قسم إلى عامل معين. ولإجراء التحليل هذا ليس ضرورياً فرض أى توزيع معين للمشاهدات.
- يحتوى جدول تحليل التباين على نفس مصادر التباين المعبر عنها فى النموذج الرياضى لا أكثر ولا أقل.
- بينما إجراء التحليل نفسه لا يحتاج إلى فرض توزيع معين إلا أنه لإجراء اختبارات المعنوية يجب أن تكون المشاهدات موزعة توزيعاً طبيعياً حتى يمكن استخدام اختبار F.

الأثر الثابت fixed effect

إذا كان لدى المجرب عدد من المعاملات ويرغب فى دراسة أثار والفروق بين هذه المعاملات بعينها دون غيرها فإنه يطلق على أثار المعاملة آثار ثابتة ولا يجوز أن تنسحب نتائج التحليل على غير تلك المعاملات بعينها. مثال ذلك إذا أراد المجرب دراسة أثر إضافة مكون معين للعليقة أو دراسة الفرق فى الدخل بين محافظة أ ومحافظة ب ... وهكذا، فإنه فى هذه الحالة ليس للمعاملة مكون تباين variance component ولكنه مجرد قيمة مربعة لأن مستويات المعاملة (أو الأقسام) لم تؤخذ بطريقة عشوائية ويرمز له بالرمز K^2 وليس σ^2 كما سيأتى فيما بعد. ويكون الاهتمام الأساسى فى هذه الحالة هو دراسة المتوسطات وليس التباين بينها.

الأثر العشوائى random effect

وفيه يختار المجرب مستويات المعاملة بطريقة عشوائية من عدة مستويات محتملة. فمثلاً إذا أراد المجرب دراسة ما إذا كان هناك فروق بين السلالات فى نموها، فإنه يختار عدداً من السلالات عشوائياً من عشيرة من السلالات ويجرى التحليل على هذه السلالات. وفى هذه الحالة يمكن للمجرب أن يعمم نتائجه على عشيرة السلالات التى أخذت منها العينة ويكون لبين السلالات مكون تباين σ^2 لأنه تباين حقيقى وليس مجرد قيمة مربعة كما هو الحال فى الأثر الثابت. وفى حالة الأثر العشوائى فإنه عادة ما يكون الهدف الأساسى من التحليل هو دراسة التباين فى عشيرة السلالات وليس متوسطات السلالات المختارة عشوائياً.

وتظل طريقة حساب مجموع المربعات ومتوسطها ودرجات الحرية واحدة فى كل من نماذج الأثر الثابت ونماذج الأثر العشوائى. الاختلافات فقط تكون فى اختبارات فرض العدم وتقدير مكونات التباين ولاسيما عندما يكون تحليل التباين متعدد الاتجاهات وليس فى اتجاه واحد. وسيتضح ذلك فى معالجات قادمة.

وإذا كانت كل العوامل فى النموذج ذات أثر ثابت يطلق عليه نموذج ثابت fixed mode (نموذج من النوع الأول model I). أما إذا كانت أثار العوامل فى النموذج عشوائية فيطلق عليه نموذج عشوائى random model (نموذج من النوع الثانى model II). وفى حالة ما إذا احتوى النموذج على بعض العوامل ذى أثر ثابت وأخرى ذى أثر عشوائى سمى بالنموذج المختلط mixed model (نموذج من النوع الثالث model III).

لاحظ أنه فى جميع أنواع النماذج السابقة يعتبر المتوسط العام ذا أثر ثابت والخطأ ذا أثر عشوائى.

مثال ٩-١

البيانات التالية تمثل الأوزان النهائية لذكور الحملان المسمنة لفترة 120 يوما بعد الفطام على خمسة علائق مختلفة. والمراد معرفة ما إذا كانت الاختلافات بين العلائق تؤدى إلى اختلاف الوزن باعتبار أن الحيوانات قد بدأت من أوزان متقاربة. (تم طرح 40 كيلوجرام من الأوزان لسهولة الحساب).

	العليقة					
	أ	ب	ج	د	هـ	
	5	9	3	2	7	
	4	7	5	3	6	
	8	8	2	4	9	
	6	6	3	1	4	
	3	9	7	4	7	
المجموع $Y_{i.}$	**26**	**39**	**20**	**14**	**33**	**132**
المتوسط $\overline{Y}_{i.}$	**5.2**	**7.8**	**4.0**	**2.8**	**6.6**	**5.28**

الحل التفصيلى:

(١) فرض العدم:

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$، أى أن المتوسطات للعلائق متساوية.

الفرض البديل:

$H_1 : \mu_i \neq \mu_i'$، أى أنه على الأقل يوجد متوسطان غير متساويين. ويحدد مستوى المعنوية، أى حجم منطقة الرفض $\alpha = 0.05$ مثلا.

(٢) النموذج الرياضى:

$$Y_{ij} = \mu + \tau_i + \in_{ij}$$

حيث:

Y_{ij} = وزن الحمل j تحت المعاملة i ، حيث $i = 1,\cdots,5$ أى 5 معاملات ، $j = 1,\cdots,5$ ، أى 5 مكررات فى كل معاملة،

μ = المتوسط العام،

τ_i = تأثير المعاملة i كانحراف عن المتوسط العام،

$\in_{ij}$ = الخطأ التجريبى $\in_{ij} \sim NID(0, \sigma^2)$، أى يتوزع طبيعياً بمتوسط 0، وتباين σ^2.

(٣) الحسابات:

(أ) مجموع المربعات الكلى المصحح: Total sum of squares (TSS)

$$TSS = \sum_i \sum_j Y_{ij}^2 - \frac{\left(\sum_i \sum_j Y_{ij}\right)^2}{tn} = \sum_i \sum_j Y_{ij}^2 - \frac{(Y_{..})^2}{tn} \qquad (٩-٣)$$

$$= 5^2 + 4^2 + ... + 7^2 - \frac{(132)^2}{25} = 834 - 696.96 = 137.04$$

وله 24 درجة حرية، والقيمة $(Y..)^2 / tn$ تعرف بمعامل التصحيح correction factor (CF)

(ب) مجموع المربعات بين المعاملات (العلائق)

Between treatments sums of squares (BSS)

$$BSS = \sum_i \frac{Y_{i.}^2}{n} - CF \qquad (٩-٤)$$

$$= \frac{26^2 + 39^2 + ... + 33^2}{5} - 696.96 = 79.44$$

وله 4 درجات حرية.

(ج) مجموع المربعات داخل المعاملات (الخطأ)

Within treatment sums of squares (WSS)

$$WSS = TSS - BSS \qquad (٩-٥)$$

بمعنى طرح مجموع المربعات بين المعاملات من مجموع المربعات الكلى المصحح أى $137.04 - 79.44 = 57.6$ ، وله 20 درجة حرية.

وعادة ما يتم تلخيص وعرض البيانات الخاصة بتحليل التباين فى جدول يعرف جدول تحليل التباين ويرمز له بالرمز ANOVA table، ويوضح جدول ٩-١ هذا التحليل.

جدول ٩-١ تحليل التباين لبيانات أوزان الحملان المسمنة على 5 علائق مختلفة

ANOVA table

مصدر التباين Source of variation (SOV)	درجات الحرية Degrees of freedom (df)	مجموع المربعات Sum of squares (SS)	متوسط المربعات Mean square (MS)	F المحسوبة Computed F
بين المعاملات Treatments	(t-1) = 4	79.44	19.86	6.9
داخل المعاملات Error	t(n-1) = 20	57.6	2.88	
الكلى المصحح C-Total	**(tn – 1) = 24**	**137.04**		

(٤) الاختبار:

بما أن القيمة المحسوبة لاختبار F

$$= \frac{\text{متوسط المربعات بين المعاملات}}{\text{متوسط المربعات داخل المعاملات}} = 6.9$$

وبمقارنتها بقيمة F من جدول F (جدول ٥ ملحق أ) التى يتم الكشف فيها عن القيمة التى تمثل هذه النسبة بين متوسطى المربعات عند مستوى المعنوية المحدد قبل التجربة α أى منطقة الرفض. إذا كانت قيمة F المحسوبة أكبر من تلك الموجودة فى الجدول، فإن فرض العدم يرفض.

ويلاحظ أن جدول F يختلف عن جدول t لأن جدول F به مدخلان، أحدهما أفقى يمثل عدد درجات الحرية بين الفئات أو المعاملات والآخر رأسى يمثل درجات حرية الخطأ.

وفى المثال السابق إذا حددت منطقة الرفض $\alpha = 0.05$ وعند درجات حرية بين المعاملات = 4 وداخل المعاملات = 20 فإن قيمة F الجدولية $F_{(4,20,0.05)} = 2.87$ ، أى أن هناك سببا لرفض فرض العدم وقبول الفرض البديل وهو أن متوسطات المعاملات مختلفة عن بعضها.

مثال ٩-٢

يمكن الحصول على نفس الحل للمثال ٩-١ باستخدام برنامج SAS كالتالى:

```
DATA LAMB;
 INPUT TRT $ WEIGHT @@;
CARDS;
A 5 A 4 A 8 A 6 A 3 B 9 B 7 B 8 B 6 B 9 C 3 C 5 C 2 C 3 C 7
D 2 D 3 D 4 D 1 D 4 E 7 E 6 E 9 E 4 E 7
PROC ANOVA;
CLASS TRT;
MODEL WEIGHT = TRT;
RUN;
```

لاحظ:

استخدمت العلامات @@ حتى يمكن قراءة البيانات بالتتابع بدلاً من وضع كل معاملة فى سطر منفصل.

استخدمت علامة $ للدلالة على أن رمز المعاملة a, b , c , d, e تقرأ كحروف هجائية وليس كأرقام.

تم وضع البيانات بحيث إن كل رقم يسبقه رمز المعاملة الخاصة به.

استخدمت proc anova للحصول على تحليل التباين، مع ملاحظة أن أثر المعاملة لابد وأن يوضع فى الـــ class.

النتائج:

The SAS System

Analysis of Variance Procedure

Dependent Variable: WEIGHT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	79.4400000	19.8600000	6.90	0.0012
Error	20	57.6000000	2.8800000		
Corrected Total	24	137.0400000			

لاحظ أنه قد تم حذف بعض نتائج البرنامج والتى سوف تذكر فيما بعد.

٩-٧ اختبار جميع الفروق بين أزواج المتوسطات (فصل المتوسطات)

Test of all differences between pairs of means -pairwise comparisons

بعد إجراء اختبار F ووجود أساس لرفض فرض العدم بعدم وجود فروق معنوية بين متوسطات المعاملات وذلك فى حالة المعاملات الثابتة، أى أن البيانات تتبع عشائر مختلفة فى متوسطاتها ولكن لها نفس التباين، وأيضا فى حالة عدم وجود مقارنات comparisons معينة مختارة مقدما لإجرائها فإنه قد يكون من المناسب إجراء اختبارات بين كل زوج من المتوسطات، وبالتالى تحديد أى المتوسطات تختلف عن بعضها وأيها لا يختلف، وهذا ما يطلق عليه بفصل المتوسطات. وقد تكون هذه المقارنات مقررة مقدما قبل بدء التجربة وتسمى المقارنات فى هذه الحالة apriori comparisons أو آجلا بعد الحصول على النتائج الفعلية وتسمى aposteriori comparisons. وهذا يعنى أنه لو أن F معنوية فإنها لا تعنى بالتالى أن جميع المتوسطات مختلفة عن بعضها البعض. فإن بعض الفروق بين المتوسطات تعتبر فروقا حقيقية فى حين أن بعض الفروق تمثل اختلافات ولا تمثل فروقا حقيقية بين العشائر واختبار F فى هذه الحالة يمثل متوسطا لمعنوية الفروق ولكن لا يوضح معنوية فروق معينة. كما قد تتصادف بعض الحالات فى أن تكون F غير معنوية ولكن يكون هناك فرق بين بعض أزواج المتوسطات.

٩-٧-١ اختبار أقل فرق معنوى The least significant difference (LSD)

إذا لم يوجد اتجاه معين لإجراء مقارنات محددة فإنه يمكن استخدام اختبار يعرف باختبار أقل فرق معنوى (LSD)، وهو بالأساس يعتبر اختبار t على مستوى معنوية محدد. ويعتبر هذا الاختبار من أسهل طرق فصل المتوسطات وأكثرها شيوعا.

وحيث إن الانحراف القياسى للفرق بين متوسطين فى العشيرة يقدر من المعادلة (٨-٥) أى $\sqrt{2\sigma^2/n}$ وحيث إن التباين فى العشيرة بين الأفراد المعاملة بنفس المعاملة وهو σ^2 يتم تقديره من تحليل التباين بواسطة S^2 أى متوسطات المربعات داخل المعاملات وعليه فان التقدير estimate للانحراف المعيارى للفرق بين متوسطى معاملتين هو $\sqrt{2\sigma^2/n}$ وأن قيمة S^2 هذه لها درجات حرية الخطأ.

وفى المثال ٩-١ وجد أن $S^2 = 2.88$ ولها 20 درجة حرية فى الخطأ وحيث إن كل متوسط يقدر من $n = 5$ مشاهدات، فإن تقدير الخطأ القياسى للفرق بين متوسطين هو $\sqrt{2\sigma^2/n} = \sqrt{2(2.88)/5} = 1.073$، وأن قيمة t الجدولية على مستوى معنوية 5% ودرجات حرية 20 يساوى 2.086، فإنه حتى يكون الفرق بين أى زوج من

المتوسطات معنويا يجب أن تزيد قيمته عن $(2.086)(1.073) = 2.24$ وهى تمثل الكمية التى يطلق عليها أقل فرق معنوى وهى:

$$LSD = t\left(\sqrt{\frac{2S^2}{n}}\right) \qquad (٩-٦)$$

وهذه القيمة تمثل أقل كمية يجب أن يصل إليها الفرق بين متوسطين حتى يقال إن هذا الفرق حقيقى (طبعا بمستوى معنوية محدد). وإذا أجرى الاختبار على مستوى 1% فقط فتصبح القيمة $(2.845)(1.073) = 3.05$، ويكون الفرق بين متوسط المعاملة (ب) وهى أعلاها ومتوسط المعاملة (د) وهى أقلها $7.8 - 2.8 = 5$ kg يعتبر فرقاً معنويا على مستوى 1%، وأيضا الفرق بين متوسطى المعاملتين هـ ، د $6.6 - 2.8 = 3.8$ kg. فى حين أن الفرق بين متوسطى المعاملتين أ، ج $5.2 - 4 = 1.2$ kg لا يمثل فرقا حقيقياً.

وعند الرغبة فى إجراء مقارنة بين كل زوجين من المتوسطات فإنه يمكن ترتيب المتوسطات ترتيبا تنازليا فى جدول، ثم يتم بعد ذلك طرح أقل المتوسطات قيمة من أعلاها ومن ثم المتوسط الذى يليه فى الترتيب التصاعدى ... وهكذا. ويتم مقارنة كل من الفروق الناتجة بالقيم المحسوبة لأقل فرق معنوى LSD وإعلان الفروق الحقيقية التى تزيد قيمتها عن قيمة الاختبار. وتسهيلا على ذلك فإنه إذا كان أحد الفروق غير معنوى فإن الفرق التالى لن يكون أيضا معنويا وعليه يقل عدد المقارنات المحسوبة.

ويمكن توضيح ذلك فى جدول ٩-٢ لمقارنة متوسطات المعاملات الخمس فى مثال ٩-١.

جدول ٩-٢ اختبار معنوية الفروق بين أزواج المتوسطات بواسطة اختبار LSD

المتوسط	الفروق (كج)			
	- هـ	- أ	- ج	- د
ب = 7.8	1.2	2.6	3.8^{**}	5.0^{**}
هـ = 6.6		1.4	2.6	3.8
أ = 5.2			1.2	2.4
ج = 4.0				1.2

ويتم مقارنة الفروق الموضحة فى الجدول بالقيم المحسوبة LSD على مستوى المعنوية المطلوب. وتبين الاختلافات المعنوية بوضع علامة (**) على الفروق المعنوية على مستوى 1% أما إذا كان المطلوب إجراء الاختبار على مستوى 5% فيمكن وضع علامة (*) على الفروق المعنوية.

وهناك اعتراضات كثيرة على استخدام اختبار LSD دون تمييز لإجراء الاختبار بين جميع أزواج المتوسطات. وذلك لأنه قد يجرى عدد أكبر من الاختبارات أكثر مما تتحمله المعلومات المتوفرة فى التجربة وبالتالى يتم الوصول إلى قرارات يعتقد أنها صحيحة فى حين قد لا يتوافر لها نفس القدر من الاحتمال للحكم الخاطئ.

فإذا كانت هناك تجربة بها 3 معاملات فإن هناك درجتى حرية للفروق بين المتوسطات فى هذه الحالة فى حين أن هناك 3 فروق ممكنة فيما بينها. وإذا كان هناك 5 معاملات فإن عدد المقارنات بين أزواج المتوسطات الممكنة يبلغ 10 فروق فى حين أنه توجد 4 درجات حرية. وفى حالة 10 معاملات فإنه يجرى 45 مقارنة فى حين أنه توجد 9 درجات حرية فقط.

ففى حالة مقارنة 3 معاملات أ، ب، ج، فإن هناك 3 مقارنات ممكنة، ولكن إذا أخذت مقارنة (أ، ب) بالإضافة إلى المقارنة (ب، ج) أى (أ – ب) + (ب – ج) والتى تعطى (أ – ب + ب + ج) أى (أ – ج) والتى يطلق عليها المقارنة الثالثة وعلى ذلك فالأخيرة غير مستقلة عن المقارنتين السابقتين وذلك لاشتراك المعاملة (ب) فى كل من المقارنتين.

ولقد أوضح Cochran and Cox (1957) أنه فى حالة إجراء المقارنة بين أعلى متوسط وأقل متوسط فى تجربة بها أكثر من معاملتين فإن الفرق بينهما إذا تم اختباره بواسطة LSD فإنه قد يعلن عن فروق معنوية فى حين أنها ليست حقيقية، أى رفض فرض العدم فى حالات أكثر من الحقيقية. ويرجع ذلك إلى أنه عند إجراء كل المقارنات فإن حجم منطقة الرفض α يصبح كبيرا وليس عند المستوى المحدد من قبل.

فمثلا إذا كان المجرب يريد إجراء اختبار بين أعلى وأقل المتوسطات على مستوى معنوية 5% فإنه فى الحقيقة يكون حجم α فى حالة 3 معاملات 13% . أما إذا زاد عدد المعاملات وأصبح 5 معاملات، فإنه يصل إلى 29% . وفى حالة 10 معاملات يصبح احتمال وجود فرق معنوى غير حقيقى 60%، أى أنه قد يكون من الأوفق فى هذه الحالة إجراء الاختبار برمى قطعة نقود وإعلان الفرق معنويا عند الحصول على الصورة وغير معنوى عند الحصول على الكتابة حيث إنه فى هذه الحالة يصبح احتمال الخطأ 50% وهو أقل من 60% المذكورة. هذا النوع من

الاختبارات التى تتم بعد الحصول على النتائج يعرف بالمقارنات التى توجهها البيانات المتحصل عليها ولذلك فهى غير سليمة وغالبا منحازة. وحيث إن اختبار LSD يعرض التجربة والمقارنات لأكبر قدر من الخطأ من النوع الأول بالنسبة لبقية الطرق المعروضة هنا وبالتالى فهو أكثر الطرق قوة فى الاختبار (Kemp, 1975).

أما إذا كانت المقارنات التى يرغب المجرب فى تقييمها قد خططت من قبل إجراء التجربة بحيث أن تكون مستقلة عن بعضها، أى أن بعضها لا يعبر عنه فى صورة دالة خطية من المقارنات الأخرى، فإن هذه المقارنات تعرف بالمقارنات المستقلة independent or orthogonal وسوف يتم توضيحها فى تفصيل لاحق.

وهذا النوع من المقارنات يمكن اختباره بواسطة LSD دون أن يؤثر زيادة عدد المعاملات فى التجربة فى زيادة احتمالات الخطأ من النوع الأول α، فمثلا إذا كانت هناك تجربة بها 6 معاملات فإن هناك 6 متوسطات هى $\overline{Y}_1, \overline{Y}_2, \overline{Y}_3, \overline{Y}_4, \overline{Y}_5, \overline{Y}_6$. فإذا قورن $(\overline{Y}_1 - \overline{Y}_2), (\overline{Y}_3 - \overline{Y}_4), (\overline{Y}_5 - \overline{Y}_6)$ مثلا، يلاحظ أن كل معاملة لم تظهر فى أكثر من مقارنة واحدة وبالتالى فإن هذه المقارنات تعتبر مستقلة عن بعضها البعض.

وقد سبق إيضاح أنه فى حالة عدم معنوية اختبار F فإنه لا ينصح بإجراء أية مقارنات بين المتوسطات إلا إذا كانت هذه المقارنات مخططة من قبل إجراء التجربة، وأن التجربة قد صممت لاستبيان هذه الفروق، عندئذ قد يصبح استعمال LSD فى المقارنات اختباراً صحيحاً. ويوضح Fisher (1935) أنه فى حالة عدم معنوية اختبار F فإنه لا يوجد أى دليل يؤيد رفض فرض العدم وأنه يجب الحذر فى إعلان فروق معنوية لبعض المقارنات. ولقد أدى ذلك إلى تعضيد الكثير من الإحصائيين عدم إجراء اختبار LSD فى حالة عدم معنوية اختبار F لتوفير قدر من الحماية ضد الخطأ من النوع الأول.

مثال ٩-٣

استخدمت 4 علائق rations فى تغذية الحملان بعد فطامها وحسب معدل الزيادة اليومى فى الوزن بالجرام خلال فترة التجربة حيث استخدم 5 حملان فى كل معاملة. ويريد القائم بالتجربة اختبار الفرق بين كل من المعاملتين 1، 4 وأيضاً المعاملتين 2 ، 3 وكانت معدلات الزيادة فى الوزن بالجرام كالتالى:

المعاملة	المكررات					المجموع	المتوسط (جم)
١	110	98	84	42	104	**438**	**87.6**
٢	122	224	60	178	126	**710**	**142.0**
٣	84	194	162	190	184	**814**	**162.8**
٤	338	274	338	170	308	**1428**	**285.6**
						3390	**169.5**

الحل:

الشخص القائم بالتجربة لديه فكرة مبدئية عن تأثير المعاملات المستخدمة فى التجربة وذلك لأنه قام بوضع فرض لاختبار المعاملات حيث إنه سوف يقارن المعاملة ٢ مع ٣ وكذلك ١ مع ٤، وذلك لأنه يعتقد أن هناك سبباً لوجود اختلافات بين كل زوج من هذه المعاملات. ولكنه فى نفس الوقت يرى تجريب المعاملات الأربع بفرض زيادة مدى التجربة وجعلها أوسع وهو ما سوف يناقش فيما بعد.

فرض العدم: $H_\circ: \mu_i = \mu_i' = \mu$

الفرض البديل: $H_1: \mu_2 \neq \mu_3$ و $\mu_1 \neq \mu_4$

النموذج الرياضى: $Y_{ij} = \mu + \tau_i + \epsilon_{ij}$

الافتراضات: $\epsilon_{ij} \sim NID(0, \sigma^2)$

الاشتراطات: $\sum \tau_i = 0$

معامل التصحيح: $C.F. = \dfrac{(\sum Y_{ij})^2}{tn} = \dfrac{(3390)^2}{20} = 574605$

مجموع المربعات الكلى المصحح:

$(110)^2 + (98)^2 + \ldots + (308)^2 - C.F. = 151175$

مجموع المربعات للعلائق:

$$\frac{(438)^2 + \ldots + (1428)^2}{5} - C.F. = 104939.8$$

مجموع المربعات الداخلى (الخطأ) $151175 - 104939.8 = 46235.2$

والجدول ٩–٣ يمثل جدول تحليل التباين

جدول ٩–٣ تحليل التباين للبيانات الخاصة بمثال ٩–٣

SOV	df	SS	MS	F - test
بين العلائق Ration	3	104939.8	34979.9**	12.1
داخل العلائق Error	16	46235.2	2889.7	
الكلى C-Total	**19**	**151175**		

وبما أن F الجدولية $F_{(3,16,0.01)} = 5.29$ فإن النتائج توضح وجود اختلافات بين المعاملات ويرفض فرض العدم وبالتالى يمكن إجراء اختبار LSD لتوضيح أى من المقارنات المخططة معنوية.

وبما أن المعاملات متساوية فى تكراراتها فإن:

الخطأ القياسى للفرق بين متوسطى معاملتين $\sqrt{\frac{2(2889.7)}{5}} = 34\ g$

وحيث إن اختبار F قد أجرى على مستوى معنوية 0.01 فإنه بالتالى سوف يتم اختبار LSD على نفس المستوى . وحيث قيمة t الجدولية = 2.921

إذا أقل فرق معنوى $LSD = (2.921)(34) = 99.3\ g$

المقارنة الأولى (عليقة ٢،٣): $162.8 - 142 = 20.8\ g$ والفرق غير معنوى.

المقارنة الثانية (عليقة ٤ ، ١): $285.6 - 87.6 = 198\ g$ والفرق معنوى (0.01) .

وقد يكون من الملائم أيضاً هنا توضيح أنه إذا افترض أنه يراد إجراء اختبار لمعنوية جميع الفروق (وهو ما لا يعتبر اختباراً جيداً فى هذه الحالة) يتضح أن المعاملة الرابعة كانت الوحيدة التى يختلف متوسطها عن بقية المعاملات. أما المعاملات الأخرى فهى غير مختلفة عن بعضها. ويمكن توضيح ذلك دون كتابة جداول لكل الفروق. وذلك بترتيب المتوسطات تنازلياً أو تصاعدياً مثلا ثم ربط المتوسطات التى لا تختلف معنوياً عن بعضها بخط أسفلها أما التى لا ترتبط ببعضها فهذا يدل على اختلافها معنوياً كما يلى:

	(١)	(٢)	(٣)	(٤)
المتوسطات	87.6	143	162.8	285.6

كما أن هناك صورة أخرى لتوضيح ذلك بأن يوضع نفس الحرف الهجائى على كل المعاملات التى لا تختلف معنوياً ووضع حرف آخر على تلك التى تختلف عن بعضها كالآتى:

	(١)	(٢)	(٣)	(٤)
المتوسطات	B 87.6	B 143	B 162.8	A 285.6

وطريقة العرض هذه تصلح فى حالة اختلاف عدد الوحدات التجريبية فى المعاملات.

وعموماً فإن اختبار LSD يعتبر اختبارا معقولا وسهلا وقوياً بالمفهوم الإحصائى وذلك لإجراء المقارنات بقيمة واحدة فقط وكما إنه فى حالة المقارنات المخططة والمستقلة فإن الاختبار يعتبر صحيحاً، حيث الاختبار يعتبر اختبار مقارنة بين متوسطين comparisonwise test.

مثال ٩–٤

استخدام برنامج SAS لإجراء اختبار LSD لبيانات مثال ٩–١

```
DATA LAMB;
 INPUT TRT $ WEIGHT @@;
CARDS;
A 5 A 4 A 8 A 6 A 3 B 9 B 7 B 8 B 6 B 9 C 3 C 5 C 2 C 3 C 7
D 2 D 3 D 4 D 1 D 4 E 7 E 6 E 9 E 4 E 7
PROC ANOVA;
CLASS TRT;
MODEL WEIGHT = TRT;
MEANS TRT/ LSD;
RUN;
```

لاحظ:

تم استخدام برنامج SAS بنفس الخطوات المذكورة فى مثال ٩–٢ مع إضافة اختيار means trt/LSD والذى يعنى طلب حساب LSD لمتوسطات المعاملات.

ولابد أن يأتى هذا الاختيار بعد الـ model وان تشتمل الـ class على الأثر المطلوب حساب LSD له.

إذا لم يذكر مستوى المعنوية فى البرنامج فإن LSD الناتجة سوف تكون عند 0.05 ويمكن إضافة 0.01 عند الرغبة فى حساب LSD عند مستوى معنوية 1% كالتالى: means trt/ LSD alpha = 0.01;

النتائج:

The SAS System

Analysis of Variance Procedure

Dependent Variable: WEIGHT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	79.4400000	19.8600000	6.90	0.0012
Error	20	57.6000000	2.8800000		
Corrected Total	24	137.0400000			

Analysis of Variance Procedure
T tests (LSD) for variable: WEIGHT

NOTE: This test controls the type I comparison wise error rate not the experiment wise error rate.

Alpha= 0.05 df= 20 MSE= 2.88
Critical Value of T= 2.09
Least Significant Difference= 2.2389

Means with the same letter are not significantly different.

T Grouping		Mean	N	TRT
A		7.800	5	b
B	A	6.600	5	e
B	C	5.200	5	a
D	C	4.000	5	c
D		2.800	5	d

بمعنى أن المتوسطات التى يصحبها نفس الحرف الهجائى لا تختلف معنوياً عن بعضها.

وهذه نفس النتائج التى سبق الحصول عليها.

٩-٧-٢ اختبار توكى Tukey's test

نظراً لما قيل عن الـ LSD ومن تعريضه التجربة والمقارنات إلى خطأ أكبر من النوع الأول، حاول Tukey أن يجد حلا لعلاج هذا باختبار أطلق عليه HSD (honestly significant difference) ويستخدم فيه جداول معدلة عن جدول (t) (جدول ١٥ ملحق أ). ويتصف اختبار توكى هذا بأنه يعطى قدرا أكبر من الحماية من خطأ النوع الأول، وعليه فهو يعرض التجربة والمقارنات إلى خطأ أكبر من النوع الثانى (أى لا يرفض فرض العدم بينما هو غير صحيح) وبالتالى فهو أقل قوة فى إعلان الفروق المعنوية إذا وجدت. وطريقة حسابه تماثل إلى حد كبير طريقة حساب LSD مع استخدام قيمة Q بدلا من t الأصلية كما تستخدم فى اختبار توكى S^2/n بدلا من $2S^2/n$ حيث دمجت $\sqrt{2}$ فى الجدول لتوفر خطوة على الحاسب وعليه تكون:

$$HSD = Q_\alpha \sqrt{\frac{S^2}{n}} \qquad (٩-٧)$$

من جدول ٩-٣ الخاص بتحليل التباين لمثال ٩-٣ $S^2 = 2889.7$ له 16 درجة حرية، وكان عدد المعاملات 4 وبكل معاملة 5 مشاهدات، ومن جدول ١٥ ملحق أ تكون $Q_\alpha = 4.05$، أى:

$$HSD = (4.05)\left(\sqrt{\frac{2889.7}{5}}\right) = 97.36$$

وواضح أن HSD عند مستوى 5% أكبر من LSD عند نفس المستوى والذى كان

$$LSD = (2.12)\left(\sqrt{\frac{(2)(2889.7)}{5}}\right) = 72.08.$$

وباختبار الفروق بين كل معاملتين طبقاً لاختبار توكى، تقارن كل من هذه الفروق بالقيمة $HSD = 97.36$. فإذا تساوى زاد الفرق المحسوب عن 97.36 كان هذا الفرق معنوياً وإذا قل عن 97.36 لا يكون معنوياً، وتكون الفروق بين متوسطات المعاملات كالتالى:

٤ – ١ = 198.0 جرام وهى تزيد عن 97.36، الفرق معنوى عند 5%
٤ – ٢ = 143.6 جرام وهى تزيد عن 97.36، الفرق معنوى عند 5%
٤ – ٣ = 122.8 جرام وهى تزيد عن 97.36، الفرق معنوى عند 5%
٣ – ١ = 75.2 جرام وهى تقل عن 97.36، الفرق غير معنوى عند 5%
٣ – ٢ = 20.8 جرام وهى تقل عن 97.36، الفرق غير معنوى عند 5%
٢ – ١ = 54.4 جرام وهى تقل عن 97.36، الفرق غير معنوى عند 5%

ويلاحظ أن الفرق (٣) – (١) = 75.2 وهو معنوى عند 5% باستخدام LSD بينما هذا الفرق غير معنوى باستخدام اختبار توكى.

٩-٧-٣ اختبار المدى المتعدد الجديد لــ (دنكن):

Duncan's new multiple range test

وجد أن LSD يحقق أكبر قدر من قوة الاختبار بينما يعرض التجربة والمقارنات إلى أكبر قدر من خطأ النوع الأول، بينما HSD أو اختبار توكى على نقيض ذلك يعطى حماية أكبر ضد خطأ النوع الأول ولكنه أقل قوة فى إعلان الفروق المعنوية. لذا فإن دنكن (Duncan, 1955) حاول إيجاد حل وسط بين الاثنين بانيا نظريته على ما يلى:

إذا أخذ الفرق بين أعلى متوسط وأقل متوسط فإن هذا الفرق يحتمل أن يحتوى على أخطاء عشوائية أكبر من أى فرقٍ آخر بين متوسطين. لذلك لابد من توفر قدر من الحماية ضد خطأ النوع الأول متناسباً مع درجة الفرق هذه، أى ترتيب المتوسطات بالنسبة لبعضها. ففى مثال ٩-٣ أعلى متوسط كان للمعاملة (٤) وأقل متوسط كان للمعاملة (١)، وعليه فالفرق بينهما 198 g يكون أكبر عرضة للخطأ العشوائى أكبر من الفرق بين معاملتين متجاورتين فى ترتيبهما مثل (٤) – (٣) = 122.8 g، (٣)-(٢) = 20.8 g ... وهكذا. وعليه فكل فرق يختبر على أساس قيمة تختلف حسب بعد المتوسطات عن بعضها.

ويكون هذا الاختبار ملائما عند الرغبة فى إجراء مقارنات بين جميع أزواج المتوسطات. ويمكن إجراء هذا الاختبار سواء كانت F معنوية أم غير معنوية.

وتتلخص طريقة اختبار دنكن فى مقارنة المدى بين متوسطين بعد الترتيب التصاعدى بقيمة تعرف بالمدى significant studentized range ويرمز له بالرمز

SSR وحيث إن قيمة $LSD = t\sqrt{\frac{2S^2}{n}}$ أى أنها تساوى $t\sqrt{2S^2_{\bar{Y}}}$، وحيث إن قيمة t تستعمل لاختبار الفرق بين متوسطين وليس المدى فيما بينهما فقد صممت قيم أخرى تشابه قيمة t وتتزايد بتزايد المدى الذى تحتويه المقارنة المختبرة، ويلاحظ أنهّ فى حالة ما إذا كان المدى يحتوى على متوسطين متجاورين فى الترتيب فإن القيمة الجديدة تتطابق مع قيمة t الجدولية.

ولقد تم ضرب القيم الجديدة فى $\sqrt{2}$ لأنه ثابت لينتج ما يعرف بقيم SSR وهى الموجودة فى جدول ١٦ ملحق أ. وهذا الجدول له مدخلان، الأفقى منها يدل على عدد المتوسطات الداخلة فى المدى الذى تم فيه المقارنة وتبدأ بالرقم 2، 3، 4 ... وهكذا. أما الرأسى فإنه يعطى عدد درجات الحرية المحسوب منها الخطأ القياسى $S_{\bar{Y}}$ من جدول تحليل التباين. وهناك قيمتان لـ SSR أحداهما خاصة بمستوى (0.05)، والأخرى خاصة بمستوى (0.01). وهنا تجدر الإشارة إلى أن هذه المستويات لا تمثل احتمال حدوث خطأ من النوع الأول α كما هو الحال فى اختبار LSD حيث إن هذا ليس اختبارا للمعنوية ولكن يسمى هذا المستوى بمستوى الحماية من حدوث الخطأ الخاص بالمقارنة التى تشمل المدى special protection level.

وبتطبيق اختبار دنكن على البيانات الخاصة بمثال ٩-٣ ومن جدول تحليل التباين الخاص به فإن قيمة الخطأ القياسى $\sqrt{2889.7/5} = 24.04$.

ولإجراء اختبار دنكن لكل أزواج المتوسطات للعلائق المختلفة يتم ترتيبها تنازلياً وأخذ الفروق بينها بنفس طريقة الجدول ٩-٢، والتى يوضحها الجدول ٩-٤.

جدول ٩-٤ اختبار الفروق بين المتوسطات بواسطة اختبار دنكن فى مثال ٩-٣

المتوسط	الفروق		
	المتوسط - (١)	**المتوسط - (٢)**	**المتوسط - (٣)**
(٤) 285.6	198.0*	143.6*	122.8*
(٣) 162.8	75.2*	20.8	-
(٢) 142.0	54.4	-	
(١) 87.6	-		

وبالتالى فعند الرغبة فى إجراء الاختبارات على مستوى 0.05 فإن قيم SSR حسب عدد المتوسطات P فى المدى وعند درجات حرية الخطأ 16 تكون:

P	2	3	4
SSR	3.00	3.15	3.23

وعلى هذا الأساس فإنه لمقارنة متوسط المعاملة (٤) بمتوسط المعاملة (١) يحتوى المدى على 4 متوسطات وتصبح قيمة اختبار دنكن:

$$(SSR)(S_{\overline{Y}}) = (3.23)(24.04) = 77.65$$

وبما أن القيمة أقل من الفرق المشاهد = 198 إذا يعلن المدى معنويا. أما فى حالة مقارنة متوسط المعاملة (٤) بمتوسط المعاملة (٢) فإن المدى يحتوى 3 متوسطات وتصبح القيمة $(3.15)(24.04) = 75.73$ ، وفى حالة مقارنة مدى يحوى متوسطين فقط تصبح القيمة $(3.0)(24.04) = 72.12$ ، ... وهكذا.

ويلاحظ هنا أن المقارنة تتم بأكثر من قيمة فى حين أنه فى حالة LSD كانت هناك قيمة واحدة لإجراء المقارنة بها.

ويلاحظ أنه فى حالة وجود متوسطين فقط فإن نتائج الاختبارات الثلاثة لابد وأن تكون متطابقة؛ وذلك لأن هناك درجة حرية واحدة بين المتوسطين، وأيضاً هناك فرق واحد دون غيره. بينما إذا وجد ثلاث معاملات بينها درجتان حرية وبينها ثلاث مقارنات هى: (أ – ب)، (أ – ج)، (ب – ج). وهنا قد يبدأ الاختلاف فى نتائج الطرق الثلاث، وكلما زاد عدد المعاملات كلما كانت هناك فرصة أكبر لأن تختلف الطرق الثلاث عن بعضها. والجدول التالى يبين عدد المعاملات وعدد المقارنات المستقلة (أى درجات الحرية) وعدد المقارنات الممكنة بين هذه المتوسطات متوسطات:

عدد المعاملات	عدد المقارنات المستقلة = (عدد المعاملات – 1) = درجات حرية المعاملات	عدد المقارنات الممكنة = $\frac{(\text{عدد المتوسطات})(\text{عدد المتوسطات} - 1)}{2}$
2	1	1
3	2	3
4	3	6
5	4	10
6	5	15
7	6	21
10	9	45

وواضح أنه كلما زاد عدد المعاملات كلما زاد الفرق والنسبة بين عدد المقارنات الممكنة ودرجات الحرية، وهذه مدعاة أكبر لأن تختلف الطرق الثلاث عن بعضها.

ويوجد طرق أخرى لفصل المتوسطات ولكل مواصفات ومزايا وعيوب. إلا إن هذه الطرق الثلاث السابق ذكرها هي أكثرها شيوعاً في التجريب الزراعي وعلى المجرب أن يقرر أيها يستخدم طبقاً لما يريده من الحماية أو السماح تجاه نوعي الخطأ. وعموماً يمكن القول إن LSD أكثرها قوة وأكثرها خطأ من النوع الأول وعلى النقيض منها اختبار توكي HSD وبينهما هو اختبار دنكن (Kemp, 1975; Gill, 1973; Balaam, 1963). وغني عن البيان أنه يمكن حساب كل من توكي و دنكن على مستوى (0.01).

مثال ٩-٥

استخدام برنامج SAS لإجراء اختبارات توكي و دنكن لبيانات مثال ٩-٣

```
DATA LAMB;
INPUT TRT GAIN @@;
CARDS;
1 110 1 98 1 84 1 42 1 104 2 122 2 224 2 60 2 178 2 126
3 84 3 194 3 162 3 190 3 184 4 338 4 274 4 338 4 170 4 308
PROC ANOVA;
CLASS TRT;
MODEL GAIN = TRT;
MEANS TRT/ TUKEY DUNCAN;
RUN;
```

لاحظ:

لم تستخدم علامة $ في input وذلك لتمييز مستويات المعاملة بالأرقام بدلا من الحروف.

استخدم كل من اختيار TUKEY و DUNCAN مع اختيار means.

النتائج المتحصل عليها عند مستوى معنوية 0.05 ويمكن تغيير ذلك إلى مستوى 0.01 كما سبق ذكره.

النتائج:

```
Analysis of Variance Procedure
Class Level Information
Class    Levels    Values

TRT        4       1 2 3 4
Number of observations in data set = 20
```

Dependent Variable: GAIN

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	104939.800	34979.933	12.11	0.0002
Error	16	46235.200	2889.700		
Corrected Total	19	151175.000			

Duncan's Multiple Range Test for variable: GAIN

NOTE: This test controls the type I comparison wise error rate, not the experiment wise error rate

Alpha= 0.05 df= 16 MSE= 2889.7

Number of Means	2	3	4
Critical Range	72.07	75.58	77.77

Means with the same letter are not significantly different.

Duncan Grouping	Mean	N	TRT
A	285.60	5	4
B	162.80	5	3
B	142.00	5	2
B	87.60	5	1

Tukey's Studentized Range (HSD) Test for variable: GAIN

NOTE: This test controls the type I experiment wise error rate, but generally has a higher type II error rate than REGWQ.

Alpha= 0.05 df= 16 MSE= 2889.7
Critical Value of Studentized Range= 4.046
Minimum Significant Difference= 97.27
Means with the same letter are not significantly different.

Tukey Grouping	Mean	N	TRT
A	285.60	5	4
B	162.80	5	3
B	142.00	5	2
B	87.60	5	1

صندوق ٩-٢

- يقصد بفصل المتوسطات هو اختبار أى من المتوسطات تختلف عن بعضها البعض.
- ومن العديد من الطرق الإحصائية المستخدمة فى فصل المتوسطات والتى تم تناولها فى هذا المؤلف:
  ١- أقل فرق معنوى LSD
  ٢- اختبار توكى HSD
  ٣- اختبار دنكن DMRT
- يعتبر LSD أكثر الطرق الثلاث إظهاراً للفروق المعنوية؛ لذا فهو أقواها ولكنه يرتكب خطأ من النوع الأول (أى رفض فرض العدم بينما يكون صحيحاً) هو الأكبر قيمة بين الثلاثة. لذا وكحماية من هذا النوع من الخطأ يفضل عدم استخدام الـــ LSD إلا إذا كان اختبار F فى تحليل التباين معنوياً.
- أما اختبار توكى فهو على النقيض من الـــ LSD لذا فهو يرتكب خطأ أكبر من النوع الثانى أى إلى عدم رفض فرض العدم مع أنه خطأ.
- أما دنكن فهو بين الاثنين ويمكن استخدامه حتى لو لم يعط اختبار F فى تحليل التباين قيماً معنوية.

٩-٨ المقارنات المستقلة (المتعامدة) Orthogonal comparisons

كما ذكر سابقاً أنه باستعمال اختبارات المعنوية سواء فى ذلك LSD أو اختبار دنكن فإن هناك عدداً من المقارنات بين المتوسطات أكثر من عدد درجات الحرية الموجودة للاختلافات بين مستويات المعاملات وبالتالى لا يمكن أن تكون هذه المقارنات مستقلة عن بعضها.

سميت المقارنات بالمتعامدة orthogonal نتيجة للمفهوم الرياضى الذى يعنى أنه فى المثلث قائم الزاوية فإن مربع الوتر = مجموع مربعى الضلعين الآخرين

(فيثاغورث). أما فى حالة المثلثات الحادة أو المنفرجة، يكون هناك زيادة أو نقص لمربع الوتر على مربعى الضلعين الآخرين، حيث الزيادة أو النقص تحددها العلاقة بين المسقط وأحد الضلعين، وهذا يعنى عدم استقلال التأثيرات. وهذا إلى حد كبير مفهوم التغاير covariance بين تأثير القوى المختلفة.

فإذا كان هناك تجربة بها عدد t من المعاملات فإنه يوجد $(t-1)$ درجات حرية للفروق بين المعاملات وأيضاً نفس العدد من المقارنات المستقلة فيما بينها أما بالنسبة للعدد الكلى الممكن للمقارنات بين أزواج المعاملات فإنه يصبح $t(t-1)/2$.

والمقارنات المستقلة تمكن من تقسيم مجموع المربعات إلى أجزاء مستقلة عن بعضها كل منها يمثل مقارنة ما لها درجة حرية واحدة. فمثلا إذا كان هناك مجموعة من المعاملات لإضافة مكون أو أكثر بكميات أو نسب مختلفة فإنه يمكن مبدئياً قياس مدى اختلاف المستويات عن بعضها ككل أو مقارنة مستوى بآخر. كما وأنه فى بعض الحالات يكون هناك أكثر من عامل واحد ولكل عامل أكثر من مستوى واحد، وفى هذه الحالة فإن المقارنات المستقلة تمكن من بيان تأثير توليفات combinations مختلفة من العوامل على الناتج النهائى. وفى حالات أخرى فإن المقارنات المستقلة تمكن من معرفة شكل العلاقة هل هى علاقة خطية linear أم أنها علاقة غير خطية non linear or curve linear.

ويتوقف تقسيم مجموع المربعات بين المعاملات إلى أجزائه على تكوين توليفة (دالة) خطية linear combination (function) من متوسطات المستويات المختلفة المراد إدخالها فى المقارنة (يفضل استخدام المجاميع بدلا من المتوسطات عند تساوى أعداد المشاهدات فى كل مستوى).

ففى المثال ٩–٣ كان هناك 4 علائق مختلفة يراد معرفة تأثيرها على معدل الزيادة اليومية فى الوزن. ولو افترض أن المعاملات الأربع تمثل عاملين مختلفين هما فيتامين معين وله مستويان (إضافة الفيتامين وعدم إضافته) أما العامل الآخر هو مضاد حيوى وله أيضاً مستويان (إضافة المضاد الحيوى وعدم إضافته) وعلى ذلك فسوف يكون هناك أربع معاملات للمستويات هى عدم إضافة كل من الفيتامين والمضاد الحيوى (معاملة ١)، إضافة الفيتامين فقط (معاملة ٢)، إضافة المضاد الحيوى فقط (معاملة ٣) و إضافة كل من الفيتامين والمضاد الحيوى (معاملة ٤).

عند الرغبة فى تقييم توليفة خطية لدراسة المقارنة بين إضافة أى من المادتين بالمقابل عدم الإضافة على الإطلاق فهذا يعنى مقارنة بين متوسط المعاملة الأولى مع متوسط المعاملات الثلاث الأخرى وتصبح التوليفة الخطية بين المتوسطات الأربعة كالتالى:

$$L=\overline{Y}_1-\left(\frac{1}{3}\overline{Y}_2+\frac{1}{3}\overline{Y}_3+\frac{1}{3}\overline{Y}_4\right)$$

ويلاحظ أن كل من المتوسطات قد دخل فى هذه التوليفة بمعامل coefficient معين، فالمعامل لمتوسط المعاملة (١) هو 1، أما المعامل لمتوسطات المعاملات الثلاث الأخرى هو $-\frac{1}{3}$. ويمكن إعادة كتابة التوليفة السابقة بضرب المعاملات فى 3 للتخلص من الكسور وتصبح:

$$L=3\overline{Y}_1-\left(\overline{Y}_2+\overline{Y}_3+\overline{Y}_4\right)$$

وبذلك تكون المعاملات $(3,-1,-1,-1)$. والصورة العامة التى يمكن بها وضع التوليفة الخطية هى:

$$L=\lambda_1\overline{Y}_1+\lambda_2\overline{Y}_2+\cdots+\lambda_t\overline{Y}_t \qquad (٩-٨)$$

حيث قيم λ_i هى المعامل الخاص بكل متوسط لمستوى $\overline{Y}_i$. وحتى تصبح التوليفة ممثلة لمقارنة تمكن من حساب مجموع مربعات خاص بها فإن مجموع المعاملات لابد من أن يساوى صفراً أى $\sum\lambda_i=0$. ففى المعادلة (٩-١٢) مجموع المعاملات يساوى صفر $\{[(3)+(-1)+(-1)+(-1)]=0\}$. ويمكن التعبير عن التوليفة فى صورة المجاميع للمستويات بدلا من المتوسطات فتصبح:

$$L=\lambda_1 T_1+\lambda_2 T_2+\ldots+\lambda_t T_t \qquad (٩-٩)$$

وينطبق على هذه المقارنة نفس ما ينطبق على المتوسطات. ويمكن حساب مجموع المربعات الخاص بأى مقارنة من خلال الدالة الخطية الخاصة بها. فإذا استخدمت المتوسطات يصبح مجموع المربعات للمقارنة:

$$\text{Sum of squares due to a comparison}=\frac{nL^2}{\sum\lambda_i^2} \qquad (٩-١٠)$$

أما إذا استخدمت المجاميع فإن مجموع المربعات الخاص بالمقارنة يصبح:

$$\text{Sum of squares due to a comparison}=\frac{L^2}{n\sum\lambda_i^2} \qquad (٩-١١)$$

وهذا ينتج طبعاً عن العلاقة بين المجموع والمتوسط.

وبالطبع من الممكن تكوين أكثر من مقارنة (دالة خطية) بين المتوسطات أو المجاميع كل منها مستقل عن الأخرى ولها مجموع مربعات مستقل. وعدد هذه المقارنات يساوى عدد درجات الحرية بين المستويات، أى (t-1) وحتى تصبح مقارنتان مستقلتين أو متعامدتين orthogonal فإنه لابد من توافر شرطين:

أ- أن مجموع المعاملات لكل من المقارنتين يساوى صفراً أى: $\sum_i \lambda_i = 0$.

ب-أن مجموع حاصل ضرب المعاملات لكل من المقارنتين يساوى صفراً أى $\sum \lambda_i \lambda_j = 0$ حيث λ_i وهو المعامل للمتوسط $\overline{Y}_i$ فى المقارنة الأولى فى حين أن λ_j هو المعامل لنفس المتوسط فى المقارنة الثانية. وبناء على هذين الشرطين، يمكن تقسيم أو تجزئة مجموع المربعات بين المستويات إلى أجزاء مستقلة أو متعامدة خاصة بمقارنة ما.

ومن ذلك يتضح أنه إذا أجريت جميع المقارنات المستقلة الممكنة بين المستويات والتى عددها $(t-1)$ وجمعت هذه الأجزاء فإنها تعطى مجموع المربعات بين المعاملات دون تداخل فيما بينها.

عند تطبيق ما سبق ذكره على المثال ٩-٣ فإنه يمكن حساب عدد من المقارنات بين المجاميع أو المتوسطات الخاصة بالمعاملات عددها يساوى عدد درجات الحرية وهو 3 = 1-4 درجات حرية أى 3 مقارنات مستقلة ممكنة. المقارنة الأولى التى سبق ذكرها وهى اختلاف مجموعة المقارنة control عن المعاملات التى أضيف فيها الفيتامين والمضاد الحيوى. وباستخدام القيم الموجودة فى هذا المثال وتطبيق المعادلة (٩-٨) فإن: $L_1 = 3(87.6) - (142 + 162.8 + 285.6) = -327.6$.

وبتطبيق المعادلة (٩-١٠) يصبح مجموع المربعات الخاص بهذه المقارنة:

$$\frac{(5)(-327.6)^2}{(3)^2 + (-1)^2 + (-1)^2 + (-1)^2} = 44717.4$$

وحيث إن مجموع المربعات للمقارنة مرتبط بدرجة حرية واحدة خاصة بها وعلى ذلك فإن متوسط المربعات لنفس المقارنة يصبح نفس القيمة (القسمة على واحد) وبالتالى يمكن اختبار معنوية هذه المقارنة بواسطة اختبار F بقسمة متوسط المربعات الخاص بها على متوسط مربعات الخطأ من جدول ٩-٣ أى $F = 44717.4 / 2889.7 = 15.5$. وهذا الاختبار يقارن بقيمة F الجدولية والتى قيمتها $F_{(1,16,0.01)} = 8.53$، وهذا يعنى أن الفرق بين نمو الحملان المعاملة والحملان فى مجموعة المقارنة control كان معنوياً.

المقارنة الثانية والتى يمكن اختبارها قد تكون الفرق بين إضافة أى من الفيتامين أو المضاد الحيوى منفردين وبين إضافتهما معا. وفى هذه الحالة يتم استبعاد مجموعة المقارنة control وتصبح المقارنة كالتالى: $L_2 = 2Y_4 - (Y_2 + Y_3)$، وقيمتها: $[(2)(1428) - (710 + 814)] = 1332$، ويكون مجموع المربعات الخاص بها:

$$\frac{(1332)^2}{(5)[(2)^2 + (-1)^2 + (-1)^2]} = 59140.8$$

يلاحظ هنا استخدام المجاميع بدلاً من المتوسطات وبالتالى تطبيق المعادلة (٩-١١) للحصول على مجموع المربعات. ويتم اختبار هذه المقارنة بنفس الطريقة أى أن: $F = 59140.8 / 2889.7 = 20.5$ وهى معنوية أيضاً عند 1%

وتطبيقاً للشروط الخاصة بكون هاتين المقارنتين مستقلتين orthogonal، يتم تكوين جدولا يوضع فيه المجاميع والمعاملات كما فى جدول ٩-٥.

جدول ٩-٥ المعاملات للمقارنات المختلفة المستقلة

	المعاملة ١	٢	٣	٤	$\Sigma \lambda_i$	$\Sigma \lambda_i^2$	$n\Sigma \lambda_i^2$	مجموع المربعات للمقارنة
المجموع	438	710	814	1428	-	-	-	
المقارنة الأولى Comparison 1	3	-1	-1	-1	0	12	60	44717.4
المقارنة الثانية Comparison 2	0	-1	-1	2	0	6	30	59140.8
المقارنة الثالثة Comparison 3	0	-1	1	0	0	2	10	1081.6
								104939.8

وكما ذكر سابقاً فإن عدد المقارنات المستقلة يساوى عدد درجات الحرية بين المستويات، وبالتالى فقد تم عمل مقارنتين. أما المقارنة الأخيرة والثالثة كما تظهر فى الجدول هى مقارنة بين إضافة كل من الفيتامين بمفرده أو إضافة المضاد الحيوى بمفرده. وقد وجد أن مجموع المربعات المحسوب لهذه المقارنة 1081.6 وعند إجراء اختبار F لهذه المقارنة $1081.6 / 2889.7 = 0.37$، أى أنها كانت غير معنوية.

ومن الواضح من جدول ٩-٦ أن مجموع المربعات للمقارنات الثلاث تجمع لمجموع مربع الانحرافات المحسوب بين المعاملات فى الجدول ٩-٣ ويرجع ذلك إلى أن كلاً منها مستقلة عن الأخرى. كما يتضح أن مجموع حاصل ضرب العوامل لكل مقارنتين يساوى صفراً فمثلا بالنسبة للمقارنتين (1, 2):

$$(3)(0)+(-1)(-1)+(-1)(-1)+(-1)(2)=0$$

وبالنسبة للمقارنتين (1, 3):

$$(3)(0)+(-1)(-1)+(-1)(1)+(-1)(0)=0$$

وهكذا ...

ويمكن إعادة كتابة جدول ٩-٣ الخاص بمثال ٩-٣ بصورة أكثر تفصيلا فى جدول ٩-٦ كما يلى:

جدول ٩-٦ جدول تحليل التباين مع الإشارة إلى مجموع المربعات للمقارنات المستقلة

SOV	df	SS	MS
TRT	3	104939.8	34979.9**
Comparison 1	1	44717.4	44717.4**
Comparison 2	1	59140.8	59140.8**
Comparison 3	1	1081.6	1081.6
Error	16	1081.6	2889.7
C - Total	**19**	**151175.0**	

ويطلق على اختبار F فى هذه الحالة لاختبار معنوية المقارنات، اختبار F المستقل orthogonal F test.

كما يمكن أيضا استخدام اختبار t المستقل orthogonal t test. ولإجراء اختبار t المستقل تستخدم المعادلة التالية:

$$t=\frac{\left|\sum\lambda_i\overline{X}_i\right|}{\sqrt{(\sum\lambda_i^2)\frac{S^2}{n}}} \qquad (٩-١٢)$$

حيث S^2 من جدول ٩-٣ تساوى 2889.72 ، $n = 5$ وبالتالى تكون t للمقارنة الأولى:

$$t = \frac{|(3)(87.6)+(-1)(142)+(-1)(162.8)+(-1)(285.6)|}{\sqrt{\left[(3)^2+(-1)^2+(-1)^2+(-1)^2\right]\frac{2889.7}{5}}} = \frac{|-327.6|}{83.28} = 3.93$$

و t للمقارنة الثانية:

$$t = \frac{|(2)(285.6) - (142+162.8)|}{\sqrt{\left[(2)^2+(-1)^2+(-1)^2\right]\frac{2889.7}{5}}} = \frac{266.4}{58.89} = 4.52$$

وهكذا بالنسبة للمقارنة الثالثة.

وتقارن قيمة t المحسوبة بقيمة t الجدولية التى قيمتها $t_{(16, 0.01)} = 2.92$ وعليه فإن الفرق للمقارنة الأولى معنوى، أى أن الفرق بين متوسط المعاملة الأولى مع متوسط المعاملات الثلاث الأخرى مؤكد وليس راجع للصدفة بدرجة ثقة 99%، وأيضاً الفرق بين إضافة أى من الفيتامين أو المضاد الحيوى وبين إضافتهما معاً أيضاً معنوى ... وهكذا.

ويلاحظ أن قيمة F للمقارنة الأولى تساوى مربع قيمة t لنفس المقارنة وكذلك العلاقة بين القيمة F وقيمة t لكل من المقارنة الثانية والمقارنة الثالثة وذلك لوجود درجة حرية واحدة فى البسط فى اختبار F حيث $t^2 = F$. ويلاحظ أيضا أن النتائج متطابقة باستخدام اختبار t أو اختبار F فى هذه الحالة فيما عدا خطأ التقريب.

ويلاحظ أنه فى اختبار t المستقل يختار المجرب مجموعة مقارنات مستقلة واحدة فقط والتى تعكس ما يريده ويفضل تصميم مثل هذه المقارنات عند بدء التجربة وليس بعد الحصول على النتائج وحساب المتوسطات حتى لا يحصل المجرب على نتائج متحيزة.

٩-٨-١ المقارنات المستقلة باستخدام برنامج SAS

يمكن استخدام برنامج SAS لعمل المقارنات المستقلة باستخدام اختيار contrast مع PROC GLM أى طريقة النموذج الخطى العام والتى سوف يتم شرحها فيما بعد. وباستخدام بيانات مثال ٩-٣ يمكن كتابة البرنامج كالتالى:

```
DATA LAMB;
INPUT TRT GAIN @@;
*---- TRT 1: CCONTROL, 2: VET, 3:AB, 4:VET AND AB ----;
*---- CONTROL: NO VET. ------------;
CARDS;
1 110 1 98 1 84 1 42 1 104 2 122 2 224 2 60 2 178 2 126
3 84 3 194 3 162 3 190 3 184 4 338 4 274 4 338 4 170 4 308
PROC GLM;
CLASS TRT;
MODEL GAIN = TRT;
CONTRAST 'CONTROL VS OTHER' TRT 3 -1 -1 -1;
*----- MEANS TRT1 VERSUS OTHER TREATMENTS-----;
CONTRAST 'VETORAB VS BOTH' TRT 0 -1 -1 2;
*----- MEANS TRT1AND TRT3 VERSUS TRT4 --------;
CONTRAST 'VET VS AB' TRT 0 -1 1 0;
*----- MEANS TRT2 VERSUS TRT3 -----------;
RUN;
```

لاحظ:

تم استخدام proc glm بدلاً من proc anova حيث إنه لا يمكن استخدام اختيار contrast مع proc anova.

يوضع اختيار contrast يتبعه اسم للمقارنة المرغوب عملها، ويمكن وضع أى اسم ذى معنى للمجرب. يلى اسم المقارنة المعاملات الخاصة بتلك المقارنة.

النتائج:

General Linear Models Procedure
Class Level Information

Class	Levels	Values
TRT	4	1 2 3 4

Number of observations in data set = 20
General Linear Models Procedure

Dependent Variable: GAIN

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
TRT	3	104939.800	34979.933	12.11	0.0002
CONTROL VS OTHER	1	44717.400	44717.400	15.47	0.0012
VETORAB VS BOTH	1	59140.800	59140.800	20.47	0.0003
VET VS AB	1	1081.600	1081.600	0.37	0.5493
Error	16	46235.200	2889.700		
Corrected Total	19	151175.000			

٩-٩ تحليل التباين عند عدم تساوى التكرارات

كثيرا لا تتساوى أعدد المكررات أو المشاهدات فى العينات أو المعاملات المختلفة فى حالة البيانات التى يتم تجميعها من مصادرها. أما فى حالة التجارب التى يخطط لها ويقوم بها المجرب فى كثير من المجالات فإن العينات تكون غالبا متساوية من حيث المكررات.

ففى حالة تساوى المعاملات فى أعداد المشاهدات بها فإن طريقة الحساب التى استخدمت للحصول على مجاميع المربعات فى فصل ٩-٦ كانت بقسمة مجموع مربعات مجاميع المعاملات على عدد المكررات داخل أى معاملة، وهو ما كان متساوياً، مطروحاً منه معامل التصحيح، وأيضا تم حساب درجات الحرية للخطأ من اعتبار أن قيمة n متساوية وبالتالى كان العدد يساوى $t(n-1)$ كما فى جدول ٩-١.

أما فى حالة عدم تساوى العينات فى مكرراتها فيصبح العدد الكلى للمشاهدات هو المجموع الكلى لها بتجميع عددها فى كل معاملة أى يصبح $\sum_i n_i$ حيث تمثل n_i عدد المشاهدات فى المعاملة i. وبما أن عدد المعاملات لا يتغير وهو t فإن عدد درجات الحرية للمجموع وداخل المعاملات يصبح $\left(\sum n_i - 1\right)$، $\left(\sum n_i - t\right)$ على الترتيب. فإذا افترض أن $N = \sum n_i$ عندئذ يصبح عدد درجات الحرية الكلى فى التجربة $N-1$ وعدد درجات الحرية داخل المعاملات (الخطأ) $N-t$.

وكما ذكر مسبقا فإن اختلاف عدد المعاملات فى تحليل التباين للبيانات أحادية التقسيم لا يؤثر على خاصية التجميع ولكن يلاحظ أن طريقة الحسابات تختلف قليلا كما يلى:

معامل التصحيح $CF = \dfrac{(Y_{..})^2}{N}$، حيث $N = \sum n_i$

$$TSS = \sum_i \sum_j Y_{ij}^2 - CF \qquad (٩-١٣)$$

وهو مجموع المربعات الكلى. وبالتالى مجموع المربعات بين المعاملات يكون:

$$BSS = \sum_i \frac{Y_{i.}^2}{n_i} - CF \qquad (٩-١٤)$$

وبالطبع وحيث لا تتأثر خاصية التجميع فى هذا النوع من التحليل يمكن الحصول على مجموع المربعات داخل المعاملات (الخطأ) WSS بالطرح كما سبق فى المعادلة (٩-٩) أو ، بطريقة مباشرة:

$$WSS = \sum_i \sum_j Y_{ij}^2 - \sum_i \frac{Y_{i.}^2}{n_i} \qquad (٩-١٥)$$

والمعادلة الأخيرة مهمة حيث توضح مفهوم الخطأ. وكثيراً ما يكون هناك احتياج إلى استخدامها عندما يراد حساب الخطأ مباشرة. ويلاحظ أن خاصية التجميع تنعدم فى الحالات الأكثر تعقيداً كما سيأتى فيما بعد. كما أن اختبار F فى هذه الحالة لا يتأثر حيث إن متوسط المربعات بين المعاملات له $(t-1)$ درجات حرية ومتوسط المربعات للخطأ له $(N-t)$ درجات حرية. أما التعديلات التى يجب الإشارة إليها فى هذا المجال فهى خاصة بإجراء اختبارات المعنوية للفروق بين كل متوسطين إما بواسطة اختبار LSD أو بواسطة اختبار دنكن وفيما يلى توضيح التعديلات المطلوبة فى كل حالة.

أولا: فى حالة اختبار LSD حيث يلزم تقدير الانحراف المعيارى للفرق بين متوسطى أى معاملتين فى التجربة وبفرض المعاملتين k, i فإن الخطأ القياسى لمقارنة متوسطى المعاملتين هو:

$$S_{(\bar{Y}_i - \bar{Y}_k)} = \sqrt{S^2\left(\frac{1}{n_i} + \frac{1}{n_k}\right)} \qquad (٩-١٦)$$

حيث S^2 هو متوسط المربعات للخطأ كالسابق و n_i ، n_k هما عدد المشاهدات أو التكرارات لكل من المعاملتين i و k على التوالى. وبالتالى تصبح قيمة $LSD = (t)S_{(\bar{Y}_i - \bar{Y}_k)}$ وتتغير هذه القيمة بتغير المتوسطات للمعاملات المقارنة أما درجات الحرية الخاصة بقيمة t فهى لا تزال درجات حرية الخطأ فى جدول تحليل التباين $(N-t)$.

ثانيا: فى حالة إجراء اختبار دنكن سبق إيضاح أن قيمة الاختبار DMRT هى $(SSR)(S_{\bar{Y}})$ وحيث إنه لا توجد قيمة ثابتة للخطأ القياسى $S_{\bar{Y}}$ وذلك لتغير التكرارات وعلى ذلك فإنه لمقارنة متوسطى معاملتين مختلفتين فى تكراراتهما تستخدم قيمة S^2 فى المعادلة. وبذلك يصبح الاختبار كالتالى:

$$DMRT = SSR\sqrt{\frac{S^2}{2}\left(\frac{1}{n_i} + \frac{1}{n_k}\right)} \qquad (٩-١٧)$$

وسبب القسمة على $\sqrt{2}$ يرجع إلى وجود الكمية داخل القيم الجدولية كما سبق إيضاحه.

مثال ٩-٦

أُجريت تجربة من 3 معاملات أ، ب، ج على الحملان وكانت أوزان الحملان بالكيلوجرام فى المعاملات الثلاث كالتالى:

المعاملة	المكررات					المتوسط
أ	32.5	32.5	33.0			**32.67**
ب	35.0	34.0	34.5	37.0	33.5	**34.80**
ج	30.5	31.5	31.0	31.5		**31.12**

جدول تحليل التباين

SOV	df	SS	MS	F
TRT	2	30.85	15.29^{**}	1.96
Error	9	8.15	0.906	
C - Total	**11**	38.73		

وعند إجراء اختبار LSD بين المتوسطات تكون المقارنات كالتالى:

قيمة اختبار LSD	المقارنة: الخطأ القياسى	المقارنة: القيمة
$(.727)(2.262) = 1.644$ الفرق غير معنوى	$\sqrt{0.906\left(\frac{1}{3}+\frac{1}{4}\right)} = 0.727$	أ - ج = 1.55
$(.695)(2.262) = 1.57$ الفرق غير معنوى	$\sqrt{0.906\left(\frac{1}{5}+\frac{1}{3}\right)} = 0.695$	ب - أ = 2.13
$(.639)(2.262) = 1.45$ الفرق معنوى	$\sqrt{0.906\left(\frac{1}{4}+\frac{1}{3}\right)} = 0.639$	ب - ج = 3.68

وعند إجراء اختبار الفروق باستخدام اختبار Duncan، ترتب الفروق بين متوسطات المعاملات تنازليا فتكون كالتالى 34.8, 32.67, 31.12 للمعاملات ب، أ، ج، على التوالى. يحوى المدى المختبر للفرق بين المعاملتين ب، ج 3 متوسطات

وبالتالى قيمة SSR عند 9 درجات حرية للخطأ تساوى 3.34 وتكون قيمة اختبار دنكن على مستوى معنوية 5% كالتالى:

$$SSR = 3.34\sqrt{\frac{0.906}{2}\left(\frac{1}{5}+\frac{1}{4}\right)} = 1.508$$

وبمقارنته بالفرق بين المتوسطين ب، ج وقدره 3.68 kg يتضح أن الفرق معنوى كما سبق إيضاحه.

ويجدر الإشارة هنا أنه فى حالة عدم تساوى عدد الوحدات التجريبية فى المعاملات ممكن أن يحدث فرق أكبر بين متوسطين فى أحدهما أو كليهما أعداد قليلة تكون غير معنوية بالمقارنة بفرق أصغر ولكن بين متوسطين فى أحدهما أو كليهما أعداد أكبر، وفى هذه الحالة لا تصلح طريقة ربط المتوسطات غير معنوية الفرق بخط واحد ولكن تستعمل الحروف للدلالة على معنوية الفروق.

٩-١٠ النموذج التجميعى الخطى The linear additive model

بعد تقديم كيفية إجراء تحليل التباين وماهيته وبعض الافتراضات الخاصة بالأخطاء errors وتوزيعها فإنه يوجد بعض الافتراضات الخاصة بتأثير المعاملات أيا كان نوعها. وكما سبق ذكره، فإن النموذج الرياضى لأى مشاهدة يتكون من تجميع ثلاثة أجزاء وهى المتوسط العام μ وانحراف متوسط المعاملة عن المتوسط العام α_i وأخيرا الخطأ الطبيعى للمشاهدة حول متوسط المجموعة المنتمية إليها. وكما سبق يوجد نوعان من الافتراضات الخاصة بأثر المعاملة وهما إما أن يكون أثر المعاملة ثابتاً fixed أو أن يكون أثرها عشوائياً random.

أولا: التأثير الثابت Fixed effect (Model I)

فى هذه الحالة يفترض أن مجموع التأثيرات $\sum\alpha_i = 0$، وأنه لو أعيدت التجربة فإنه يتوقع الحصول على نفس المجموعة من التأثيرات فى التجربة الجديدة. ويكون النموذج الرياضى بالتالى وكما سبق ذكره فى (٩-٦) هو:

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij} \qquad (١٨-٩)$$

ثانيا: التأثير العشوائى Random effect (Model II)

إذا كان تأثير المعاملات a_i عشوائياً فإن ذلك يعنى أن المستويات التى فى التجربة تعتبر عينة مأخوذة من عشيرة كبيرة لا نهائية من تلك المستويات، وأن متوسط هذه العشيرة هو صفر، حيث إنها أيضا تعتبر انحرافات عن المتوسط العام μ.

وفى هذه الحالة فإنه عند إعادة التجربة يتم الحصول على عينة جديدة من عشيرة المستويات هذه.

وبما أن مجموعة المستويات فى هذا النموذج مأخوذة عشوائياً من عشيرة المستويات والذى متوسطه يساوى صفرا فإن المستويات أيضاً لها تباين σ_a^2 وهو ما يراد أيضا تقدير قيمته من تحليل نتائج التجربة.

بمعنى أنه فى هذه الحالة لا يكون الاهتمام بإيجاد اختبار لمدى معنوية التأثيرات لمجموعة محددة من المستويات كما فى النموذج الأول، ولكن الاهتمام يكون أكثر بتقدير مدى ما تشارك به هذه المستويات فى التباين الكلى حيث إنها عنصر من عناصر التباين بين المشاهدات المختلفة. وعلى هذا الأساس يصبح متوسط المربعات mean squares عبارة عن جزئين، الأول راجع للتباينات الناتجة من عشيرة المستويات المستخدمة فى التجربة σ_a^2 والثانى هو التباين الراجع لاختلاف المشاهدات داخل المستوى الواحد من المعاملات وهى σ_e^2 أو الخطأ التجريبى.

وبذا يصبح التباين بين مشاهدتين مأخوذتين عشوائيا من نفس المستوى (المعاملة) يرجع إلى التباين الداخلى σ_e^2 (الخطأ) فى حين أن التباين بين مشاهدتين مأخوذتين عشوائيا من مستويين مختلفين يرجع إلى كلا التباينين أى $\sigma_e^2 + \sigma_a^2$. وبذلك فإنه فى حالة النموذج العشوائى Model II فإنه يفترض أن الخطأ $\in_{ij}$ يتوزع طبيعيا بمتوسط صفر وتباين σ_e^2 ويرمز لذلك $\in \sim NID(0, \sigma_e^2)$ وأيضا فان مستويات المعاملات a_i تتوزع طبيعيا بمتوسط صفر وتباين σ_a^2 ويرمز لذلك $a \sim NID(0, \sigma_a^2)$. ويكون النموذج فى هذه الحالة:

$$Y_{ij} = \mu + a_i + \in_{ij} \qquad (١٩-٩)$$

وهناك أوجه للشبه والاختلاف بين كل من النموذجين يمكن تلخيصها فيما يلى:

أ- فى النموذج I يفترض أن α_i تأثيرها ثابت وبالتالى فإن $\sum \alpha_i = 0$، وكذلك $\in \sim NID(0, \sigma_e^2)$، ويراد تقدير قيمتها واختبار معنوية الفروق بين هذه المجموعة المحددة من المستويات.

ب- فى النموذج II يفترض أن a_i تأثيرها عشوائى وبالتالى $a \sim NID(0, \sigma_a^2)$ وكذلك $\in \sim NID(0, \sigma_e^2)$، وبالتالى فإن هناك اهتماماً واضحا تجاه تقدير حجم σ_a^2.

ج- اختبار فرض عدم $H_0 : \mu_i = \mu$ يتماثل مع اختبار الفرض بأن $H_0 : \sigma_a^2 = 0$ حيث إن فى كلتا الحالتين α_i أو a_i لابد أن تساوى صفراً، وأن اختبار F يعتبر صحيحاً فى كلتا الحالتين لفرض العدم.

د- متوسط المربعات S^2 داخل المعاملات يعتبر تقديراً لقيمة σ_e^2 فى العشيرة، فى حين أنه يمكن إثبات أنه فى حالة اختلاف المتوسطات μ_i فإن متوسط المربعات بين المعاملات فى النموذج I يعتبر تقديراً غير منحاز للكمية $\sigma_e^2 + n\sum\alpha_i^2/(t-1)$، حيث كما سبق $\alpha_i = \mu_i - \mu$، أى أنه بالإضافة إلى التباين العادى بين الأفراد المعاملة بنفس المعاملة فهناك تأثير اختلاف متوسطات المعاملات والذى يتم تربيعه ثم يحسب متوسطه بالقسمة على درجات الحرية $(t-1)$ وبالطبع فإن الكمية α_i هى كمية ثابتة فى هذا النموذج. أما بالنسبة للنموذج II فإن متوسط المربعات للخطأ هو نفسه فى حالة النموذج I، أما متوسط المربعات بين المعاملات فهو تقدير للكمية $\sigma_e^2 + n\,\sigma_a^2$ حيث σ_a^2 هى مكون التباين بين المعاملات. وحيث n تمثل عدد الأفراد أو المشاهدات داخل كل مستوى فإنه يمكن حساب تقدير σ_a^2 من S_a^2 من تحليل التباين كالتالى:

مكون التباين بين المعاملات:

$$S_a^2 = \frac{\text{MS between} - \text{MS within}}{n} \qquad (٩-٢٠)$$

وبالتعويض بالقيم المتوقعة لكل من متوسطى المربعات تكون معادلة التوقع:

$$S_a^2 = \frac{(\sigma_e^2 + \sigma_a^2) - \sigma_e^2}{n} \cong \sigma_a^2 \qquad (٩-٢١)$$

وبالتالى فإن S_a^2 المحسوبة من التحليل تعتبر تقديراً لقيمة σ_a^2 لمكون التباين بين المجموعات.

وفى المثال المذكور لتحليل التباين فى جدول ٩-١ وبافتراض أن الاختلافات بين العلائق المختلفة تأثيرها عشوائى، أى النموذج الرياضى II، فإن التقدير المتحصل عليه لتباين الخطأ $\sigma_e^2 = 2.88$، متوسط المربعات بين المعاملات 19.86 وهو تقدير للقيمة $\sigma_e^2 + n\,\sigma_a^2$ وحيث $n = 5$، عدد الحملان داخل كل عليقة، وعلى ذلك فإن تقدير مكون التباين الراجع لأثر العليقة $3.396 = \frac{19.86 - 2.88}{5} = \sigma_a^2$. أما إذا كان تأثير المعاملات ثابتاً (نموذج I) فإن توقع متوسط المربعات يكون $\sigma_e^2 + nK_a^2$.

٩-١١ تقدير مكونات التباين فى حالة اختلاف حجم العينات

فى كثير من الحالات والتى يكون التعامل فيها مع بيانات فى مجالات خاصة مثل تربية الحيوان أو الدواجن أو الأسماك أو فى دراسة النواحى الوراثية والاجتماعية فى حالة المجتمعات الإنسانية، أو الدراسات الوراثية فى النبات، يكون حجم العائلات (أى عدد أفراد العائلة) فى معظمها مختلفا من عائلة إلى أخرى وهذه العائلات التى تدخل فى العينة تعتبر جزءاً عشوائياً من عشيرة كبيرة من تلك العائلات. وبالتالى فإن الاختلافات فيما بينها سواء كانت وراثية أو اجتماعية تعتبر نموذجاً لما تعني به التأثيرات العشوائية، أى النموذج الرياضى II. والذى يمكن التعبير عنه رياضياً كما يلى:

$$Y_{ij} = \mu + f_i + \epsilon_{ij} \quad , i = 1,2,...,t \quad j = 1,2,...,n$$

وبالتالى تكون افتراضات تحليل التباين كما سبق $f \sim NID(0,\sigma_f^2)$، $\epsilon \sim NID(0,\sigma_e^2)$ حيث تمثل f_i فى هذه الحالة الاختلافات العشوائية الموجودة بين العائلات سواء كانت وراثية كنتيجة لتأثير العوامل الوراثية الموروثة من الأب (الطلوقة) كما فى دراسات تربية الحيوان، أو أنها تأثيرات اجتماعية للعائلة ككل نتيجة لعوامل تعليمية أو ثقافية أو اقتصادية مختلفة كما فى الدراسات الإنسانية، وقد تكون نتيجة لاختلاف التراكيب الوراثية للخطوط المختلفة lines كما فى دراسات تربية النباتات.

وحيث إن حجم العائلات أو الخطوط يختلف من حالة إلى أخرى بحيث إن عدد المشاهدات المدروسة فى العائلة f_i هو n_i وبالتالى فإن حجم التجربة كلها $N = \sum n_i$ ولا يؤثر ذلك على تحليل التباين كما سبق إيضاحه فى حالة لنموذج I. ولكن الاختلاف هنا هو أن متوسط المربعات بين المجموعات (العائلات) يعتبر تقديراً غير منحاز للكمية $\sigma_e^2 + n_\circ \sigma_f^2$ حيث $n_\circ$ فى هذه الحالة هى المتوسط التوافقى لعدد أفراد المجموعة الواحدة وهو يساوى:

$$n_\circ = \frac{1}{t-1}\left(N - \frac{\sum n_i^2}{N}\right) \qquad (٩-٢٢)$$

وبناء على ذلك فإنه يمكن تقدير قيمة مكون التباين بين المجموعات كما سبق باستخدام المعادلة (٩-٢٠) مع استبدال قيمة n فى المقام بقيمة $n_\circ$ المحسوبة من المعادلة (٩-٢٢) كالتالى:

$$S_a^2 = \frac{\text{MS between} - \text{MS within}}{n_\circ} \qquad (٩-٢٣)$$

ويعتبر التقدير المتحصل عليه لمكون التباين بين المجموعات σ_a^2 تقديراً غير منحاز إلا أنه فى حالة عدم تساوى العينات لا يعتبر بنفس الكفاءة كما هو الحال فى حالة التساوى. ولذا فإنه ينصح، إذا كان بالإمكان، استخدام عينات ذات أحجام متساوية وذلك لسهولة معرفة التوزيع الرياضى لها فى هذه الحالة.

مثال ٩-٧

الجدول التالى يوضح إنتاج البيض الناتج لمجاميع من الدجاج معبراً عنه كنسبة مئوية حيث كل مجموعة تمثل بنات ديك واحد أى أنصاف شقيقات (قد لا تعتبر النسب المئوية مثالاً جيداً لتحليل التباين نظراً لأنه عند عدم تساوى العينات فإن اعتبار أن σ_e^2 ثابتة أو موحدة لكل المعاملات قد لا يكون صحيحاً تماماً إلا أنه نظراً لأن أغلب النسب تقع فى المدى من 25-75% فإن الفروق تكون ضئيلة، كما سيتضح ذلك عند دراسة التحويرات transformation فيما بعد).

رقم الديك	النسبة المئوية لإنتاج البيض (للبنات)							عدد البنات
١	59.7	60.0	70.5	66.4				4
٢	76.2	69.3	66.6	73.5				4
٣	62.5	71.3	52.3	65.0	73.5			5
٤	75.6	58.0	71.0	49.0	69.0			5
٥	64.3	57.3	44.3	62.5	59.5	50.0		6
٦	62.7	50.0	69.6	34.0	69.0	52.3		6
٧	45.0	58.3	69.1	72.9	77.8	75.2		6
٨	71.8	58.4	24.7	55.7	50.0	68.3	57.5	7
٩	66.8	46.3	50.5	68.0				4
إجمالى								47

وبالطبع فإن الاختلاف بين ما يورثه الآباء للأبناء يعتبر كميات عشوائية مأخوذة من عشيرة كبيرة من هذه الاختلافات الوراثية وبالتالى فإنه باستخدام النموذج الثانى II:

$$Y_{ij} = \mu + s_i + \in_{ij} \quad , i = 1, 2, ..., 9 \quad j = 1, 2, ..., n_i$$

حيث تعتبر s_i انحراف وراثة الأب i عن المتوسط العام μ، $\in_{ij}$ هو الاختلافات الطبيعية بين الأفراد والتى ترجع إلى أسباب وراثية وأخرى بيئية. وكلا الكميتين عشوائيتان أى $s \sim NID(0,\sigma_s^2)$، $\in \sim NID(0,\sigma_e^2)$ ويجري التحليل الإحصائى لهذه البيانات للحصول على المعلومات التالية التى يوضحها جدول تحليل التباين التالى:

ANOVA table

SOV	df	SS	MS	EMS
Between parents	8	1320.65	165.04	$\sigma_e^2 + n_o\sigma_s^2$
Within parents	38	4685.39	123.30	σ_e^2
C - Total	**46**	**6006.04**		

EMS: Expected mean squares

وحيث إنه يكون هناك اهتمام فى هذه الحالة بتقدير إلى أى مدى توجد تباينات راجعة إلى تأثير اختلاف الآباء فإن ذلك يستوجب الحصول على تقدير لمكون التباين σ_s^2 الخاص بتأثير الأب وبالتالى يجب حساب n_o طبقا للمعادلة (٩-٢٧):

$$n_o = \frac{1}{8}\left[47 - \frac{(4^2 + 4^2 + ... + 7^2 + 4^2)}{47}\right] = 5.197$$

وهذه لا تختلف كثيراً عن قيمة $\overline{n} = 5.22$ ولكنها أدق. وبالتعويض فى (٩-٢٣) يمكن الحصول على تقدير σ_s^2 كالتالى:

$$\sigma_s^2 = \frac{165.04 - 123.3}{5.197} = 8.024$$

علماً بأن التقدير المحسوب لمكون التباين داخل المجموعات هو نفسه متوسط المربعات داخل المجموعات، أى: $\sigma_e^2 = 123.3$.

مثال ٩-٨

استخدام برنامج SAS لحل بيانات مثال ٩-٧ والحصول على مكونات التباين.

```
DATA EX10111;
INPUT SIRE EGG @@;
CARDS;
1 59.7 1 60 1 70.5 1 66.4 2 76.2 2 69.3 2 66.6 2 73.5
3 62.5 3 71.3 3 52.3 3 65 3 73.5 4 75.6 4 58 4 71 4 49 4 69 5 64.3
5 57.3 5 44.3 5 62.5 5 59.5 5 50 6 62.7 6 50 6 69.6 6 34 6 69 6 52.3
7 45 7 58.3 7 69.1 7 72.9 7 77.8 7 75.2 8 71.8 8 58.4 8 24.7 8 55.7
8 50 8 68.3 8 57.5 9 66.8 9 46.3 9 50.5 9 68
PROC VARCOMP METHOD = TYPE1;
CLASS SIRE;
MODEL EGG = SIRE;
RUN;
```

لاحظ:

استخدمت proc varcomp للحصول على مكونات التباين للمتغيرات العشوائية فى النموذج، فى هذه الحالة مكون الأب والخطأ التجريبى.

استخدمت method = type 1، وهى التى تحدد طريقة حساب مجموع المربعات حيث يوجد طرق أخرى مثل MIVQUE0 و ML (maximum-likelihood). شرح هذه الطرق والاختلافات بينها يقع خارج نطاق هذا المؤلف.

النتائج:

Variance Components Estimation Procedure
Class Level Information

Class	Levels	Values
SIRE	9	1 2 3 4 5 6 7 8 9

Number of observations in data set = 47

Dependent Variable: EGG

Source	DF	Type I SS	Type I MS
SIRE	8	1320.65378723	165.08172340
Error	38	4685.73600000	123.30884211
Corrected Total	46	6006.38978723	

Source	Expected Mean Square
SIRE	Var(Error) + 5.1968 Var(SIRE)
Error	Var(Error)

Variance Component	Estimate
Var(SIRE)	8.03817982
Var(Error)	123.30884211

٩-١٢ معامل الارتباط الداخلى (الجوانى) Intra-class correlation

فى تحليل التباين أحادى التقسيم حيث يوجد عدد من الأفراد فى كل فئة من الفئات وعندما يكون مكون التباين بين المجموعات أكبر من الصفر $(\sigma_a^2 > 0)$ فإن هذا يدل على أن أفراد المجموعة أو الفئة الواحدة تتشابه مع بعضها. وفى النموذج II لتحليل التباين فإن عدم ثبوت فرض العدم الخاص بأن $\sigma_a^2 = 0$ يوضح أن هناك سبباً للتشابه بين أفراد المجموعة الواحدة أكثر من أفراد لا تنتمى لنفس المجموعة.

وقد افترض أن كل قيم المشاهدات Y_{ij} لها نفس المتوسط μ ونفس التباين σ_e^2، ولكن أى فردين من نفس الفئة i يكون بينهما معامل ارتباط معين وهذا المعامل يطلق عليه اسم معامل الارتباط الداخلى (الجوانى) Intra – class correlation وهو يقيس مدى التشابه بين الأفراد المنتمية إلى فئة معينة لسبب ما ولكن دون أن تكون هناك علاقة سببية (صفة معينة تؤدى إلى أن يميز أحد الأفراد بوضع يختلف عن الأفراد الآخرى).

فمثلاً لا يوجد سبب عند قياس التشابه بين الأخوة الأشقاء حيث لا يؤثر أحدهما على الآخر كما هو الحال فى حالات أخرى مثل التشابه بين الأب وابنه حيث يمرر الأب نصف مورثاته إلى النسل مباشرة.

ويمكن تقدير معامل الارتباط الداخلى بالقيمة r_I من جدول تحليل التباين حيث إنه يمكن جبرياً إثبات أن متوسط المربعات بين الفئات BMS هو تقدير للكمية:

$$\sigma^2[1+(n-1)\rho_I] \qquad (٩-٢٤)$$

وأن متوسط المربعات داخل الفئات WMS فى هذه الحالة هو تقدير للكمية:

$$\sigma^2(1-\rho_I) \qquad (٩-٢٥)$$

حيث n عدد الأفراد داخل الفئة وأن σ^2 هو التباين للأفراد و ρ هو معامل الارتباط الجوانى بين الأفراد وبالتالى فإن (BMS − WMS) تعطى تقديراً للكمية $n\sigma^2\rho_I$ ومن هذا تتضح العلاقة:

$$r_I = \frac{BMS - WMS}{BMS - (n-1)(WMS)} \qquad (٩-٢٦)$$

وبالتعويض عن مكونات كل من متوسطات المربعات تعطى العلاقة:

$$r_I = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_a^2 + \hat{\sigma}_e^2} \qquad (٩-٢٧)$$

وتوقعاً فإن الحد الأدنى لمعامل الارتباط الداخلى هو صفر ولكن فى الواقع قد تنشأ القيم السالبة لمعامل الارتباط الداخلى من وجود قيمة أكبر لمتوسط المربعات داخل الفئات أكبر من متوسط المربعات بين الفئات وقد يحدث ذلك فى بعض الحالات، مثلاً فى الأرانب سينشأ تنافس بين أفراد العش الواحد أو البطن الواحدة وسوف تدفع باستمرار الأفراد الكبيرة الأفراد الصغيرة عن حصولها على الغذاء وبذلك ينتج أن التباين فى الأوزان داخل المجموعات أكبر من التباين بين المجموعات مما قد يعطى قيماً سالبة لقيمة r_I.

٩-١٣ العينات داخل العينات Sample within samples

يمكن تصور وضع تقسم فيه كل عينة إلى تحت عينات sub-samples ويمكن أيضاً لهذه الأخيرة أن تقسم إلى تحت تحت عينات sub-subsamples ... وهكذا. وبذلك تنشأ تقسيمات داخل تقسيمات مما يعطى توزيعاً متفرعاً أو عنقودياً hierarchal or nested. أو قد يكون العامل الرئيسى أو التقسيم الأكبر هو عامل ذو تأثير ثابت fixed كما سبق. فى حين أن العوامل التى فى التقسيمات التالية له تكون مختارة عشوائياً random، وبذلك ينشأ ما يعرف بالنموذج الخليط أى أن بعض التأثيرات ثابتة والأخرى عشوائية. أو قد تكون أعداد المكررات (المشاهدات) داخل العينات أو تحت العينات غير متساوية مما يؤدى إلى الحصول على بيانات غير متزنة.

٩-١٣-١ التقسيم العنقودى: حالة تمام العشوائية للعوامل

Nested or hierarchal classifications

كما سبق إذا قسمت كل عينة إلى تحت عينات sub-samples وإلى تحت تحت عينات sub-subsamples أو تقسيمات داخل تقسيمات فإن ذلك سوف يعطى توزيعاً

متفرعاً أو عنقودياً hierarchal or nested. فلو افترض مثلاً أنه توجد مجموعة من الحظائر يوضع فى كل منها ديك واحد ومعه مجموعة من الدجاج (الأمهات) وتعطى كل أم عدداً من البيض الذى يفقس ليعطى الكتاكيت (النسل) فإن أى كتكوت يمكن أن يوضع فى عينة تمثل نسل إحدى الأمهات وينشأ أيضا عن ذلك مجاميع من النسل كل منها تنتمى لديك معين. أو قد تؤخذ مجموعة من العينات من أكثر من حقل أو موقع ثم يجرى على كل عينة من العينات المأخوذة من الحقل أكثر من تقدير واحد للمكونات المطلوب الكشف عنها مثل نسبة الملوحة أو الكالسيوم... الخ. وعلى هذا الأساس فقد يكون عدد العينات من كل حقل متساوياً كما أن عدد التقديرات لكل عينة متساوياً. فى حين فى المثال الخاص بالدجاج قد يختلف عدد الأمهات لكل ديك ويختلف أيضاً عدد النسل الناتج من كل أم.

فى مثل هذه التجارب العنقودية فإن بعض العوامل تكون محتواة داخل عوامل أخرى كتحت فئات sub classes وفى هذه الحالة لا يمكن بالتالى إجراء مقارنات لتحت الفئات على كل مستويات الفئات classes.

مثال ٩-٩

لوحظ فى حديقة ما فى منطقة صناعية ظهور بقع مرضية على أوراق الأشجار وهنا قد تختلف الأشجار عن بعضها كما أن الأوراق قد تختلف أيضاً عن بعضها، ولذا فإنه يجرى عد لعدد البقع على كل من نصفى أوراق مأخوذة عشوائياً من أشجار مختارة أيضاً عشوائياً من الحديقة لتقدير حجم الإصابة مثلاً. فإذا افترض أخذ 3 شجرات ثم أخذ 4 أوراق من كل شجرة وتم عد لعدد البقع على كل من نصفى الورقة، فيكون شكل التوزيع كالتالى:

	الشجرة الأولى				الشجرة الثانية				الشجرة الثالثة			
الأوراق	١	٢	٣	٤	١	٢	٣	٤	١	٢	٣	٤
أنصاف الأوراق	١ ٢	١ ٢	١ ٢	١ ٢	١ ٢	١ ٢	١ ٢	١ ٢	١ ٢	١ ٢	١ ٢	١ ٢

وبذا يصبح حجم التجربة الكلى فى هذه الحالة عبارة عن عدد أنصاف الأوراق التى تم عليها العد (المقياس أو المشاهدة) وهى $24 = (2)(4)(3)$ نصف ورقة. ويوضح الجدول التالى القيم المشاهدة لعدد البقع فى هذا المثال:

		الشجرة الأولى				الشجرة الثانية				الشجرة الثالثة			
الورقة		١	٢	٣	٤	١	٢	٣	٤	١	٢	٣	٤
نصف الورقة	١	11	5	6	4	15	9	11	12	17	10	10	8
	٢	9	5	4	2	17	8	10	11	16	11	9	10
المجموع للورقة		**20**	**10**	**10**	**6**	**32**	**17**	**21**	**23**	**33**	**21**	**19**	**18**
المجموع للشجرة		**46**				**93**				**91**			
المجموع الكلى		**230**											

والنموذج الرياضى للتقسيم العنقودى (عاملين أو تقسيمين):

$$Y_{ijk} = \mu + \tau_i + \ell_{ij} + \epsilon_{ijk} \qquad (٩-٢٨)$$

حيث μ تمثل المتوسط العام و τ_i تأثير اختلاف الأشجار عن بعضها وقد يرجع ذلك إلى اختلافات فى المكان فى الحقل أو اختلافات مصدرها وراثى مثلاً، ℓ_{ij} هى اختلاف الأوراق عن بعضها والذى يكون نتيجة لمواقع غير محددة أو لأسباب عشوائية أخرى و ε_{ijk} وهو الخطأ الطبيعى والذى لا يؤول إلى أى من العوامل فى النموذج وهو الذى يستخدم فى اختبار فرض العدم. وفى المثال السابق i تمثل الأشجار حيث $i=1, 2, 3$ و j تمثل الأوراق حيث $j = 1, 2, 3, 4$ وأنصاف الأوراق يمثلها k حيث $k = 1, 2$. أما افتراضات النموذج فهى كتلك الخاصة بالنموذج العشوائى كنتيجة لوضع التجربة وبذلك يكون $\tau \sim N(0,\sigma^2_\tau)$، $\ell \sim N(0,\sigma^2_\ell)$، $\in \sim N(0,\sigma^2_e)$ أى أن النموذج تام العشوائية فى تأثير العوامل الداخلة فيه.

ولإجراء التحليل تستخدم المجاميع الموجودة أسفل جدول البيانات فى حساب مجاميع المربعات حسب النموذج الرياضى المفترض ويلاحظ هنا أنه ليست هناك مجاميع للصفوف حيث إن أنصاف الأوراق لا تمثل عاملاً من عوامل الاختلافات ولكن هى قياسات عشوائية على أقل وحدات عينية، أى أن الوحدة التجريبية experimental unit هى فى هذه الحالة الورقة الكاملة، أما نصف الورقة فتعتبر وحدة تعيين sampling unit يجرى عد البقع على كل من نصفى الورقة لاختبار مدى كفاءة عملية العد ودقة القياس. ويجرى التحليل كالتالى:

معامل التصحيح:

$$CF = \frac{(Y_{...})^2}{(i)(j)(k)} = \frac{(230)^2}{(3)(4)(2)} = 2204.17$$

مجموع المربعات الكلى:

$$\sum\sum\sum Y_{ijk}^2 - CF = 16^2 + 14^2 + \cdots + 10^2 - 2204.17 = 375.83$$

مجموع المربعات بين الأشجار:

$$\sum_i \frac{Y_{i..}}{(j)(k)} - CF = \frac{(46)^2 + (93)^2 + (91)^2}{(4)(2)} - 2204.17 = 176.58$$

مجموع المربعات بين الأوراق:

$$\frac{\sum\sum Y_{ij.}^2}{(k)} - CF = \frac{(20)^2 + (10)^2 + \ldots + (18)^2}{2} - 2204.17 = 362.83$$

ومجموع المربعات الأخير هذا أى 362.83 يشمل كلا من الاختلافات بين الأشجار والتى كان مجموع مربعاتها 176.58 بالإضافة إلى الاختلافات بين الأوراق داخل الشجرة الواحدة، ويمكن حساب مجموع المربعات بين الأوراق داخل الأشجار (وتكتب بين الأوراق/الأشجار) كالتالى:

$$\frac{\sum\sum Y_{ij.}^2}{(k)} - \frac{\sum Y_{i..}^2}{(j)(k)} = 362.83 - 176.58 = 186.25$$

ويمكن حساب مجموع المربعات للعينات، أى بين العينات داخل الأوراق داخل الأشجار (وتكتب بين العينات/الأوراق/الأشجار) كالتالى:

$$\sum\sum\sum Y_{ijk}^2 - \frac{\sum\sum Y_{ij.}^2}{(j)(k)} = 375.83 - 362.83 = 13$$

ويمكن تلخيص النتائج المتحصل عليها فى جدول تحليل التباين التالى مع العلم بأن التباين فى الفئة الأكبر تحتوى إضافة إلى مصدر التباين الخاص بها التباينات الناشئة من الفئات الأقل منها وبالتالى فإنه يمكن أيضاً كتابة متوسط المربعات المقدر أمام كل متوسط مربعات محسوب.

ANOVA table

SOV	df	SS	MS	EMS
بين الأشجار	2	176.58	88.290	$\sigma_e^2 + 2\sigma_\ell^2 + 8\sigma_\tau^2$
بين الأوراق/الأشجار	9	186.25	20.694	$\sigma_e^2 + 2\sigma_\ell^2$
بين العينات/الأوراق/الأشجار	12	13.00	1.083	σ_e^2
الكلى المصحح	**23**	**375.83**		

ومن الجدول يمكن حساب مكونات التباين لكل من التأثيرات الداخلة فى النموذج وتبعاً لمتوسط المربعات المقدر.

فتقدير مكون التباين للأوراق عبارة عن:

$$\sigma_\ell^2 = \frac{\sigma_e^2 + 2\sigma_\ell^2 - \sigma_e^2}{2} = \frac{20.694 - 1.083}{2} = 9.81$$

فى حين أن تقدير مكون التباين بين الأشجار عبارة عن:

$$\sigma_\tau^2 = \frac{\sigma_e^2 + 2\sigma_\ell^2 + 8\sigma_\tau^2 - \sigma_e^2 + 2\sigma_\ell^2}{8} = \frac{88.29 - 20.694}{8} = 8.45$$

ولاختبار الفرضيات الخاصة بمكونات التباين فإنه يستخدم اختبارF بحيث يكون متوسط المربعات المستخدم فى المقام يحتوى على كل ما يتوقع وجوده فى متوسط المربعات للبسط ما عدا الجزء المرغوب فى اختباره، أى أنه لاختبار معنوية σ_ℓ^2 فإن $\frac{\sigma_e^2 + 2\sigma_\ell^2}{\sigma_e^2}$، وعلى ذلك فتحت فرض العدم الخاص بأن $\sigma_\ell^2 = 0$ تصبح القيمة النظرية لاختبار $F = 1$ أى $F = \frac{20.694}{1.083} = 19.1$ ، وأيضاً بالنسبة لاختبار الفرض الخاص بأن $\sigma_\tau^2 = 0$ فإن $F = \frac{88.29}{20.694} = 4.266$. وبالتالى فإنه بالكشف عن قيمة F الجدولية $F_{(0.01,9,2)} = 4.39$ فى الحالة الأولى، يتضح أن قيمتها معنوية جداً أى أن الاختلافات بين الأوراق داخل الأشجار كانت معنوية ويرفض الفرض الخاص بأن $\sigma_\ell^2 = 0$.

وأيضاً تكون الاختلافات بين الأوراق معنوية عند مستوى معنوية 5% وبالتالى فإن الخطأ القياسى لمتوسط المعاملة (الأشجار) $S_{\overline{Y}} = \sqrt{\frac{20.694}{8}} = 1.608$ والخطأ القياسى للفرق بين متوسطى معاملتين هو $\sqrt{\frac{2(20.694)}{8}} = 2.274$.

٩-١٣-٢ حالة النموذج الخليط Mixed model

يكون العامل الرئيسى أو التقسيم الأكبر فى كثير من الحالات هو عاملا ذا تأثير ثابت fixed كما سبق. فى حين أن العوامل التى فى التقسيمات التالية له تكون مختارة عشوائياً random، وبذلك ينشأ ما يعرف بالنموذج الخليط أى أن بعض التأثيرات ثابتة والأخرى عشوائية. فلو افترض فى الدواجن مثلا وجود عدد من الحظائر، وبكل حظيرة مجموعة من الأمهات المختارة عشوائياً وأن القياسات تتم على النسل الناتج، فإن تأثير الحظائر فى هذه الحالة يكون ثابتاً حيث إن هذه هى كل المستويات المستخدمة فى التجربة فى حين أن الأمهات تأثيرها عشوائى نظراً لاختيار التزاوجات بهذه الطريقة وأيضاً النسل الناتج منها تكون الاختلافات فيما بينها عشوائية. ويكون شكل النموذج الرياضى فى هذه الحالة كالتالى:

$$Y_{ijk} = \mu + \alpha_i + d_{ij} + \in_{ijk} \qquad (٩-٢٩)$$

حيث تمثل α_i التأثير الثابت للحظائر وينطبق عليها الاشتراط $\sum\alpha_i = 0$. فى حين أن σ_d^2 تمثل التأثير العشوائى للأمهات داخل الحظائر وهى تتبع التوزيع الطبيعى بمتوسط مقداره صفر وتباين σ_d^2. بينما تمثل $\in_{ijk}$ التأثير العشوائى للاختلافات بين النسل الناتج وهو الخطأ الحقيقى والذى يتوزع طبيعيا بمتوسط صفر وتباين σ_e^2.

خطوات حساب مجاميع المربعات وخلافه هى نفسها الخاصة بالنموذج السابق إلا أنه فى العمود الخاص بمتوسط المربعات المتوقع للعامل الثابت α_i تستبدل القيمة σ_a^2 بالقيمة $\frac{\sum\alpha_i^2}{(a-1)}$ حيث $(a-1)$ تمثل عدد درجات الحرية للتقسيم الثابت.

وفى النموذج الخليط فإن متوسط أى فئة من الفئات الخاصة بالتأثير الثابت فى التقسيم الأعلى عبارة عن:

$$\overline{Y}_{i..} = \mu + \overline{\alpha}_i + \overline{d}_{i.} + \overline{\in}_{i..} \qquad (٩-٣٠)$$

حيث تمثل $\overline{d}_{i.}$ و $\overline{\in}_{i..}$ متوسط التأثيرات العشوائية الداخلة فى الفئة i. وعلى هذا الأساس فإن التباين الخاص بأى متوسط $\overline{Y}_{i..}$ يكون كالتالى:

$$V(\bar{Y}_{i..}) = \frac{\sigma_d^2}{d} + \frac{\sigma^2}{nd} = \frac{1}{nd}(\sigma^2 + n\sigma_d^2) \qquad (٩-٣١)$$

وبذا فإن متوسط المربعات للأمهات داخل (الحظائر) هو التقدير غير المتحيز للكمية $\sigma^2 + n\sigma_d^2$. وبالتالى فإن الخطأ القياسى عبارة عن الجذر التربيعى للتباين المعطى فى (٩-٣١).

٩-١٣-٣ حالة عدم تساوى حجم العينة Unequal numbers

فى تجارب وبيانات الإنتاج الحيوانى وغيره قلما أن تتوافر أعداد الحيوانات أو المشاهدات المتساوية سواء فى العينات أو العينات داخل العينات. فهناك اختلافات فى عدد التزاوجات لكل طلوقة واختلافات فى عدد الأبناء الناتجة من كل أم (فى الأرانب أو الدواجن على سبيل المثال) وبذلك يمكن القول بأن بيانات الإنتاج الحيوانى غالباً ما تكون غير متزنة.

ولقد استنبطت الوسائل الحسابية لتحليل التباين فى حالة عدم تساوى حجم العينات وتحت العينات كما وأن الطرق التى يمكن بها حساب مكونات التباين أيضاً فى تلك الحالات قد استخرجت أيضاً، ولو أنها إلى حد ما أكثر صعوبة عنها فى حالات تساوى حجم العينات وتحت العينات.

أما بالنسبة لطريقة الحساب فهى لا تختلف كثيراً عن الطريقة التى اتبعت فى تحليل التباين فى فصل ٩-١١.

مثال ٩-١٠

فى تجربة على نمو الكتاكيت بالجرام (مطروحاً منه 100 جرام) والناتجة من تزاوج ديكين مختارين عشوائياً مع مجموعة من الإناث فى فترة الأسبوعين الأولين تم الحصول على النتائج التالية:

الأمهات	١	٢	٣	٤	٥	٦	٧	٨
	الديك أ			الديك ب				
النمو فى النسل	7	10	8	6	14	8	12	16
		8	4	8	9	6	18	12
		5		8		4	20	14
						8		11
								12
								15
المجموع للأم	**7**	**23**	**12**	**22**	**23**	**26**	**50**	**80**
المجموع للديك		**42**				**201**		
العدد		**6**				**18**		

النموذج الرياضى:

$$Y_{ijk} = \mu + s_i + d_{ij} + \in_{ijk} \qquad (٩-٣٢)$$

وقد افترض أن جميع العوامل فى النموذج عشوائية فيما عدا المتوسط وكذلك $s \sim N(0,\sigma_s^2)$ و $d \sim N(0,\sigma_d^2)$ و $\in \sim N(0,\sigma_e^2)$.

العدد الكلى للنسل: $N = 24$

المجوع الكلى: $\sum_i \sum_j \sum_k Y_{ijk} = 243\,g$

مجموع المربعات الكلى: $\sum_i \sum_j \sum_k Y_{ijk}^2 = 2893$

مجموع المربعات المصحح: $\sum_i \sum_j \sum_k Y_{ijk}^2 - CF = 2893 - 2460.375 = 432.625$

مجموع المربعات بين الآباء: $\sum_i \frac{Y_{i..}^2}{n_i} - CF = 2538.5 - 2460.375 = 78.125$

مجموع المربعات بين الأمهات داخل الآباء:

$$\sum_i \sum_j \frac{Y_{ij}^2}{n_{ij}} - \sum_i \frac{Y_{i..}^2}{n_i} = 2792.167 - 2538.5 = 253.667$$

أو
$$3371.792 - 78.125 = 253.667$$

مجموع المربعات بين النسل داخل الأمهات داخل الآباء:

$$\sum_i \sum_j \sum_k Y_{ijk}^2 - \sum_i \sum_j \frac{Y_{ij.}^2}{n_{ij}} = 2893 - 2792.167 = 100.833$$

أو
$$2893 - 2460.375 - (78.125 + 253.667) = 100.833$$

وبناء عليه يمكن وضع النتائج المتحصل عليها من التحليل فى جدول تحليل التباين التالى:

SOV	df	SS	MS	EMS
Sires	1	78.13	78.12	$\sigma_e^2 + k_2\sigma_d^2 + k_3\sigma_s^2$
Dams/sires	6	253.67	42.28	$\sigma_e^2 + k_1\sigma_d^2$
Offspring /Dams/ Sires	16	100.73	6.30	σ_e^2
C-Total	**23**	**432.62**		

ويلاحظ من جدول تحليل التباين وحساب مجاميع المربعات السابقة أن مجموع المربعات بين الأمهات داخل الطلائق أى الأمهات التى تتزاوج مع أحد الطلائق يساوى مجموع المربعات بين الأمهات بدون النظر إلى مجاميع الطلائق مطروحاً منه مجموع المربعات بين الطلائق. وأيضاً درجات الحرية فهناك 8 أمهات عدد درجات الحرية فيما بينها 7 وبالتالى درجة الحرية للأمهات داخل الطلائق $7 - 1 = 6$ حيث إن هناك درجة حرية واحدة بين الطلائق. ولكن بناء على أن عدد الأفراد (المشاهدات) فى كل مستوى غير متساوى، أى أن عدد النسل لكل أم غير متساوى، وأيضاً عدد الأمهات التى تزاوجت مع كل أب وبالتالى فإن حجم عائلة كل طلوقة متغير. ويجب حساب قيم k_1, k_2, k_3 التى تظهر فى العمود الأخير من جدول تحليل التباين حتى يتسنى تقدير مكونات التباين. ولقد تم عرض الحالة التى يختلف فيها عدد الأفراد فى التقسيم الأحادى فى الفصل ٩-١١ حيث تم حساب الكمية n_o على أساس المتوسط

التوافقى لعدد الأفراد فى الفئة. ولقد أعطى Gaylor and Hartwell (1969) القيم المتوقعة لمتوسطات المربعات فى حالة التحليل العنقودى وكيفية تقدير مكونات التباين. وعموماً فإن القيم المطلوبة لمتوسط الأعداد هى:

$$k_1 = \frac{1}{df(dams)}\left(N... - \sum_i \frac{\sum_j n_{ij}^2}{n_{i.}}\right) \qquad (٩-٣٣)$$

$$k_2 = \frac{1}{df(sires)}\left(\sum_i \frac{\sum_j n_{ij}^2}{n_{i.}} - \frac{\sum_i \sum_j n_{ij}^2}{N..}\right) \qquad (٩-٣٤)$$

$$k_3 = \frac{1}{df(sires)}\left(N... - \frac{\sum_i n_{i.}^2}{N...}\right) \qquad (٩-٣٥)$$

وبتطبيق هذه المعادلات على المثال ٩-١٠ فإن حسابها كالتالى:

$$k_1 = \left(\frac{1}{6}\right)\left[24 - \left(\frac{1^2+3^2+2^2}{6} + \frac{3^2+2^2+4^2+3^2+6^2}{18}\right)\right] = 2.9259$$

$$k_2 = \left(\frac{1}{1}\right)\left[\begin{array}{l}\left(\frac{1^2+3^2+2^2}{6} + \frac{3^2+2^2+4^2+3^2+6^2}{18}\right) \\ -\left(\frac{1^2+3^2+...+6^2}{24}\right)\end{array}\right] = 2.7778$$

$$k_3 = \left(\frac{1}{1}\right)\left[24 - \frac{6^2+18^2}{24}\right] = 9$$

وبالتالى فيمكن تقدير مكونات التباين حسب جدول تحليل التباين والمعادلات التالية:

$$\sigma_d^2 = \frac{MSD - MSW}{k_1} \qquad (٩-٣٦)$$

وبالتعويض

$$\sigma_d^2 = \frac{42.278 - 6.6302}{2.9859} = 12.2957$$

أما تقدير قيمة σ_S^2 فيمكن حسابه عن طريق التعويض فى المعادلة التالية:

$$\sigma_s^2 = [MS_s - MS_w - k_1/k_2(MS_d - MS_w)]/k_3 \qquad (٩-٣٧)$$

وبالتعويض

$$\sigma_s^2 = \frac{1}{9}\left[78.125 - 6.302\left(\frac{2.778}{2.9259}\right)(42.278 - 6.302)\right] = \frac{37.6682}{9} = 4.1851$$

هذه التقديرات لمكونات التباين تعتبر غير متحيزة. أما فى حالة النموذج الذى يكون فيه العامل الأساسى (الطلائق فى هذه الحالة) ذات تأثير ثابت fixed فإن اختبار F لفرض العدم الخاص بالطلائق يكون $78.125/42.278 = 18.5$ لا يمكن أن يستخدم نظراً لاختلاف القيم المحسوبة لكل من k_1 وk_2 ولكن يشكل بسطا ومقاما لهذا الاختبار.

هذا ويمكن أن يمتد مثل هذا النوع من التحليلات إلى مستويات أو عوامل تدخل ضمن التحليل العنقودى كأن تكون مثلاً الطلائق ممثلة داخل مزارع مختلفة ... وهكذا.

مثال ٩-١١

يمكن استخدام برنامج SAS لحل مثال ٩-١٠ كالتالى:

```
DATA NESTED;
INPUT SIRE $ DAM PROG @@;
CARDS;
A 1 7 A 2 10 A 2 8 A 2 5 A 3 8 A 3 4
B 4 6 B 4 8 B 4 8 B 5 14 B 5 9 B 6 8
B 6 6 B 6 4 B 6 8 B 7 12 B 7 18 B 7 20
B 8 16 B 8 12 B 8 14 B 8 11 B 8 12 B 8 15
```

```
PROC VARCOMP METHOD = TYPE1;
CLASS SIRE;
MODEL PROG  = SIRE DAM(SIRE);
RUN;
```

لاحظ الطريقة التى كتب بها الـ model حيث توضح أن الأمهات موزعة داخل الطلائق dam(sire) وهذا هو سبب أن أرقام الأمهات للطلوقة الأول تختلف عن أرقام الأمهات للطلوقة الثانى.

نتائج التحليل:

Variance Components Estimation Procedure
Class Level Information

Class	Levels	Values
SIRE	2	A B
DAM	8	1 2 3 4 5 6 7 8

Number of observations in data set = 24

Dependent Variable: PROG

Source	DF	Type I SS	Type I MS
SIRE	1	78.12500000	78.12500000
DAM(SIRE)	6	253.66666667	42.27777778
Error	16	100.83333333	6.30208333
Corrected Total	23	432.62500000	

Source	Expected Mean Square
SIRE	Var(Error) + 2.7778 Var(DAM(SIRE)) + 9 Var(SIRE)
DAM(SIRE)	Var(Error) + 2.9259 Var(DAM(SIRE))
Error	Var(Error)

Variance Component	Estimate
Var(SIRE)	4.18541960
Var(DAM(SIRE))	12.29549051
Var(Error)	6.30208333

صندوق ٩-٣

افتراض عشوائية أو ثبات المؤثرات لا يؤثر على طريقة تقسيم درجات الحرية أو حساب مجموع المربعات أو المتوسطات ولكن يؤثر فقط علي محتويات متوسط المربعات المتوقع (EMS) وبالتالى على اختبار الفروض فى تحليل التباين.

تمارين الباب التاسع

٩-١ ثلاثة أصناف من البطاطس تمت زراعتها فى حقل متماثل مقسم إلى 12 قطعة وكانت النتائج المتحصل عليها لوزن المحصول الناتج كالتالى:

أ- 18, 20, 14, 11

ب- 18, 28, 17, 22

ج- 9, 18, 6, 7

المطلوب: اختبار الفرض الخاص بتساوى المحصول فى الأصناف الثلاثة. وما هى الافتراضات والاشتراطات المطلوبة لإجراء التحليل؟ اذكر النموذج الرياضى المستخدم.

٩-٢ العينات التالية تمثل المحصول الناتج من استخدام خمسة أنواع من الأسمدة. والمطلوب: اختبار فرض تساوى المتوسطات فى المحصول واختبار أيها تختلف عن الأخرى بواسطة استخدام اختبار LSD.

أ	ب	ج	د	هـ
21	35	45	32	45
35	12	60	53	29
32	27	33	29	31
28	41	36	42	22
14	19	31	40	36
47	23	40	23	29
25	31	43	35	42
48	20	48	42	30

٩-٣ فى تجربة ما فى الدواجن استخدم فيها 9 ديوك حيث تزاوج كل من الديوك مع 4 أمهات وقيس الإنتاج لعدد 5 أبناء من كل تزاوج مأخوذة عشوائياً. وكان متوسط الإنتاج للبيض فى التجربة 62.75 بيضة/فرد. وكان مجموع المربعات الكلى غير المصحح $\sum Y^2 = 773905.75$. استكمل بناء على ذلك جدول تحليل التباين التالى وقدر مكون التباين لكل من الديوك والأمهات والخطأ.

SOV	df	SS	MS
Sires		4732.0	
Dams/sires			397.5
Progeny/Dams/ Sires			

٩-٤ أعطيت النموذج الرياضى $Y_{ij} = \mu + \tau_i + \in_{ij}$ حيث $i = 1,\cdots,4$ و $j = 1,2$
وكان متوسط المعاملات هو 8 ,5 ,3 ,2 للمعاملات من 1 إلى 4 على التوالى،
وكان مجموع المربعات الكلى غير المصحح = 212
المطلوب:

١- إجراء تحليل التباين واختبار فرض العدم بتساوى المعاملات.

٢- إجراء المقارنات وتقسيم التباين الخاص بالمقارنة.

(أ) المعاملة 1 ضد بقية المعاملات

(ب) المعاملة 3 ضد المعاملة4

(ت) المعاملة 2 ضد المعاملتين 3، 4.

٩-٥ من جدول تحليل التباين التالى المطلوب تقدير قيمة الارتباط الداخلى إذا علمت أن قيمة $n_\circ = 5.197$

SOV	df	SS
Sires	8	1273.3
Within sires		
C-Total	**46**	**5872.1**

٩-٦ فيما يلى تقدير محتوى الفوسفور فى أربعة أصناف من أشجار الفاكهة

الصنف				
١	0.40	0.35	0.47	0.50
٢	0.70	0.65	0.79	0.84
٣	0.80	0.60	0.66	0.73
٤	0.75	0.57	0.70	0.80

أ- اختبر فرض العدم بعدم وجود فروق معنوية بين الأصناف الأربعة عند مستوى معنوية 5% .

ب- اختبر فرض العدم بأن متوسط الصنف الأول لا يختلف معنوياً عن متوسط الأصناف الأخرى.

٩-٧ جربت معاملتان غذائيتان لمعرفة تأثيرهما على وزن العرف فى الديوك بالجرام. وكانت البيانات كما يلى (علماً بأن الديوك كانت متماثلة إلى حد كبير فى أوزانها وأعمارها وحالتها الصحية):

معاملة (أ)	129	115	72	125	127	68	130	125
معاملة (ب)	70	71	79	55	75	74	90	

والمطلوب اختبار فرض العدم بتساوى متوسطى المعاملتين باستخدام

أ- اختبار t

ب- اختبار F ، تحقق من أن $F = t^2$

٩-٨ جرب نوعان من المبيدات على كمية المحصول الناتج من 30 فدان فى كل وكان متوسط إنتاج الفدان والانحراف المعيارى كما يلى (علماً بأنه قد تركت 30 فدان أخرى كمجموعة ضابطة control):

	العدد	المتوسط	الانحراف المعيارى للعينة
مبيد (١)	30	91	7.5
مبيد (٢)	30	88	8.2
مجموعة ضابطة	30	75	8.8

والمطلوب:

١- كتابة النموذج الرياضى مع النص على الافتراضات

٢- اختبر فروض العدم التالية عند مستوى 5% :

$$H_\circ : \mu_1 = \mu_2 = \mu_3$$

$$H_\circ : (\mu_1 + \mu_2)/2 = \mu_3$$

$$H_\circ : \mu_1 = \mu_2$$