

**Assumptions of analysis of variance**

- ١- عشوائية المعاينة
- ٢- استقلالية المشاهدات
- ٣- تجانس التباين
- ٤- طبيعية التوزيع
- ٥- التجمعية
- ٦- تحوير البيانات



كما سبق إيضاحه في البابين التاسع والحادي عشر وكما سيتضح فيما بعد في الباب الخامس عشر أنه لإجراء تحليل التباين وتكوين الجدول الخاص به فإن الافتراض الوحيد المطلوب هو أن تكون درجات الحرية بكل مصدر مستقلة خطياً أي linearly independent وهذا ما حققه الشرط الخاص بأن مجموع تأثير المعاملات يساوى صفراً مثلاً، ولهذا الشرط تفقد درجة حرية واحدة، ولكن الأمر يختلف عندما يراد اختبار أى فرض أو فروض خاصة بالعينة المجرى عليها التجربة. والتي عادة ما يستخدم فيها جداول  $t$ ،  $F$  أو  $\chi^2$  ... الخ. فلإجراء اختبار هذه الفروض باستخدام تلك الجداول لابد أن يستوثق المجرى أن هناك عدة شروط تتوافر في المادة التجريبية. وأهم هذه الشروط هي: عشوائية المعاينة، استقلال المشاهدات، تجانس التباينات، طبيعية التوزيع، والتجمعية. وسوف يتم في هذا الباب شرح كل من هذه الاشتراطات، كيفية الاختبار لها، كيفية معالجة البيانات إذا لم تتوافر إذا أمكن، وما الذى يمكن أن يحدث لنتائج اختبار الفروض إذا لم تتوافر مثل هذه الافتراضات.

## ١-١٢ عشوائية المعاينة Random sampling

هذا شرط أساسى لصحة اختبار أى فرض. فالعينة التى سوف يجرى عليها اختبار الفرض لابد أن تكون مأخوذة عشوائياً من عشيرة محددة التعريف والذى سوف يرد الاستنباط عليها. فإذا أراد المجرى أن يختبر أداء حملان سلالة ما على أربع علائق مثلاً يجب أولاً أن تكون هذه الحملان تمثل عينة عشوائية من حملان هذه السلالة ثم يجب أن توزع هذه الحملان عشوائياً على المعاملات الأربع دون تدخل من المجرى، وأحسن وسيلة لإجراء هذا هى استخدام جداول الأرقام العشوائية (جدول ١ ملحق أ) أو كتابة رقم كل حمل على ورقة ثم خلط هذه الأوراق جيداً ثم سحب كل مجموعة عشوائياً من مجموعة الأوراق أو من خلال الحاسب. ودون هذا الإجراء قد يتسرب إلى التجربة بعض العوامل التى تسبب غياب العشوائية عن غير قصد من المجرى، افترض أن المجرى سيسحب من الحظيرة أول مجموعة من الحملان ليعطيها المعاملة الأولى، وثانى مجموعة ليعطيها المعاملة الثانية ... وهكذا حتى المجموعة المتبقية يعطيها للمعاملة الرابعة. فى مثل هذا الإجراء قد يتسرب كما سبق القول بعض العوامل المحددة للتعشية الكاملة حيث قد تكون المجموعة الأولى من الحملان هى الأهدأ طبعاً وربما الأضعف جسماً وهكذا بالترديج فى المجاميع حتى تكون المجموعة الأخيرة التى عينت للمعاملة الرابعة هى أكثر الحملان نشاطاً (حيث لم يتمكن من الإمساك بها أولاً) وربما الأكثر حيوية ونشاطاً، وفى هذا تحيز للنتائج، فالفرق بين المعاملات ربما ستحتوى أيضاً على فرق فى الحملان نفسها وهذا طبعاً لم يقصده المجرى. حتى وإن كان ذلك "السيناريو" المذكور مبالغ فيه ولكن لن يكلف المجرى جهداً كبيراً إذا هو أجرى التعشية بالطرق سائلة الذكر.

أيضاً كثيراً ما يتم وضع الحيوانات المولودة، في محطة تجارب مثلاً، في قوائم حسب تاريخ ميلادها مثلاً، فإذا أخذ المجرّب المجموعة الأولى للمعاملة الأولى والثانية لثانية ... وهكذا، فإنه يدخل عامل العمر بطريقة غير مقصودة، وهذا أيضاً قد يؤدي إلى تحيز نتائج التجربة. وهكذا يمكن سرد عديد من الأمثل المتوقعة وغير المتوقعة التي قد تؤدي إلى تحيز النتائج دون قصد من المجرّب. وصمام الأمان في هذا هو إجراء التعشبية إجراءً سليماً، وليس للتعشبية العامة اختبار معين، ولكن يمكن اختبار مثلاً إذا كانت المعاملات موزعة عشوائياً على عامل معين مثلاً كالوزن أو العمر ... الخ. وعدم توافر هذا الشرط يؤدي إلى حالة يكون المجرّب غير واثق فيها من إرجاع الظواهر إلى مسبباتها الحقيقية والتي تحيز النتائج.

## ١٢-٢ استقلالية المشاهدات Independence of observations

المقصود هنا باستقلالية المشاهدات هو استقلال الخطأ في هذه المشاهدات أي  $e_{ij}$  في النموذج الإحصائي مثلاً  $Y_{ij} = \mu + a_i + e_{ij}$  ، فهذه  $e_{ij}$  كلها لجميع المشاهدات يجب أن تكون مستقلة عن بعضها ومن توزيعات متطابقة. فإذا فرض أن المجرّب لديه أربعة حيوانات في كل معاملة وكل معاملة وضعها في حظيرة، وفي كل معاملة يوضع كمية ثابتة من العليقة، والمطلوب اختبار الفرق بين العلائق. ففي هذه الحالة يمكن أن ينشأ وضع هو أنه إذا استهلكت بعض الحيوانات في المعاملة الواحدة نصيباً أكبر من العليقة، فإن البعض الآخر سيكون نصيبه أقل بالضرورة وينشأ عن هذا ارتباطاً سالباً بين قيم الانحرافات الـ  $e_{ij}$ 's لهذه المجموعة، وفي ذلك حيدة عن الاستقلال وستكون هذه الحيدة عن استقلال الـ  $e_{ij}$  بصورة أكبر إذا ما كانت مساحات أماكن الغذاء (المداد) محدودة الأمر الذي سيؤدي إلى تصارع (تتقاطع) الحيوانات والأقوى هو الذي سيفوز بنصيب أكبر من العليقة على حساب الأضعف وهو ما يطلق عليه *bullying effect*. ولهذا يجب أن يفتن المجرّب لمثل هذه الأمور سبباً، وأن يحاول علاجها مسبقاً طبقاً لإمكانياته كأن يعطي لكل حيوان كمية من لعليقة مستقلة عن بقية الحيوانات إذا سمح المكان بهذا أو يكثر من المساحات لمخصصة للتغذية "الطوايل" أو تقدم العليقة مجزأة على عدة مرات إذا سمحت غراض التجربة بذلك أو قص قرون الحيوانات حتى تقلل من هذا التأثير.

وهناك نوع من ترابط الأخطاء دأب كثير من باحثي الإنتاج الحيواني على تجاهله وهو الارتباط الموجب بين المشاهدات على مستوى الوقت وذلك عندما تؤخذ المشاهدة (الوحدة التجريبية) على نفس الحيوانات على فترات زمنية. فعدم أخذ هذا الإجراء في لحساب عند التخطيط أو التحليل يؤدي إلى ارتباط بين الأخطاء. فالحيوان الذي يقاس بنفسه عدة مرات في الشتاء ثم في الربيع ثم في الصيف ثم في الخريف، إذا كان نفسه

سريعاً بالنسبة لبقية الحيوانات في أحد هذه المواسم، يحتمل أن يكون سريعاً نسبياً أيضاً في بقية المواسم الأخرى وفي هذا ارتباط بين الأخطاء. ومن هذا أيضاً ظاهرة المعاومة في الأشجار والتي ينشأ عنها ارتباط سالب ... الخ، وسيعالج هذا الموقف في الباب الخامس عشر. وقد ينشأ الارتباط أيضاً في حالة عدم التعشية السليمة. ويمكن أخذ فكرة عن هذا الارتباط من حساب  $e_{ij}$  لكل مشاهدة فهذه القيم يجب ألا تأخذ اتجاهاً معيناً في القيمة أو الإشارة أو التكرار. فمثلاً لا يجب أن تكون هناك عدة قيم سالبة تتبعها أخرى موجبة، أو تكون قيمها متدرجة إلى أعلى أو إلى أسفل إذا رتبنا المشاهدات عشوائياً أو تكون منتظمة التكرار ويتوالى السالب منها والموجب في نظام معين. وهناك اختبارات للكشف عن استقلالية الأخطاء منها اختبار ديرين واطسون Dubrin/Watson statistic. وليس هناك من أسلوب لمعالجة البيانات حتى يتوافر فيها شرط الاستقلال هذا، ولكن الذي يجب عمله هو التعشية السليمة مع اختيار التصميم الإحصائي الملائم وعدم توافر شرط الاستقلالية يؤثر تأثيراً ملحوظاً على اختبار  $t$  و  $F$ .

### ١٢-٣ تجانس التباين Homogeneity of variance

والمقصود بالتباين هنا هو تباين الخطأ  $e$  في النموذج الإحصائي، ففي الأبواب السادس والتاسع والعاشر فإن تجانس التباين كان شرط مسبق لإجراء اختبارات الفروض. فإذا كان الهدف هو اختبار الفرق بين معاملتين مثلاً طبقاً للنموذج  $Y_{ij} = \mu + a_j + e_{ij}$  فالفرض هنا يعني أنه إذا قدر التباين داخل المعاملة الأولى بقيمة  $\sigma_{e1}^2$  وداخل المعاملة الثانية بقيمة  $\sigma_{e2}^2$  فإن هذين التقديرين هما تقدير لقيمة مشتركة واحدة هي  $\sigma_e^2$  أي أن الأخطاء في المعاملة الأولى وتلك التي في المعاملة الثانية يتبعان (أو مأخوذتان من) نفس العشيرة.

وهناك عدة طرق لاختبار فرض تجانس التباينات أي

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 = \sigma_e^2$$

### ١٢-٣-١ اختبار بارتلت لتجانس التباين

#### Bartlett's (1937) test for homogeneity of variance

في هذا الاختبار تقدر قيمة  $B$  كما يلي:

$$B = \frac{2.3026}{C} \{ [\sum (n_i - 1)] \log_{10} \bar{S}^2 - \sum (n_i - 1) \log_{10} S_i^2 \} \quad (1-12)$$

حيث

$$C = 1 + \frac{1}{3(k-1)} \left[ \sum \frac{1}{n_i - 1} - \frac{1}{\sum(n_i - 1)} \right]$$

$S_i^2$  : تقدير غير متحيز لتباين ( $\sigma_i^2$ ) المعاملة  $i$

$\bar{S}^2$  : متوسط التباين

$\log_{10}$ : اللوغاريتم للأساس 10، وإذا استخدم اللوغاريتم الطبيعي الطبيعي natural logarithm (الأساس  $e$ ) فإن عامل الضرب 2.3026 لا يستخدم في المعادلة (١٢-١)

$n_i$  : عدد المشاهدات في المعاملة  $i$

$K$  : عدد التباينات

والإحصاء  $B$  يكون موزعاً حسب  $\chi^2$  بدرجات حرية  $(k-1)$ .

### مثال ١٢-١

إذا كانت أوزان الحملان في ثلاث مجموعات كالتالي:

المجموعة الأولى: 20، 15.2، 19، 13.1، 14، 19.1

المجموعة الثانية: 20.7، 21.7، 16.9، 15.9، 16.2، 21

المجموعة الثالثة: 15.9، 16.7، 13.6، 22.8، 20.6، 17

اختبر الفرض أن التباينات الثلاث لهذه المجموعات متساوية أي:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$$

يمكن إجراء اختبار بارنلت لتجانس التباين كالتالي:

$1/(n_i - 1)$	$(n_i - 1) \times \log_{10} S_i^2$	$\log_{10} S^2$	متوسط المربعات $S^2$	درجات الحرية $(n_i - 1)$	مجموع المربعات SS	المجموعة
0.2	4.7437	0.9487	8.8867	5	44.4333	الأولى
0.2	4.2632	0.8527	7.1227	5	35.6133	الثانية
0.2	5.2435	1.0487	11.1867	5	55.9333	الثالثة
<b>0.6</b>	<b>14.2504</b>			<b>15</b>	<b>135.9799</b>	<b>المجموع</b>

$$\bar{S}^2 = \sum (n_i - 1)\sigma_i^2 = \text{TSS}/N = 135.9799/15 = 9.0653$$

$$C = 1 + \frac{1}{(3)(3-1)} \left[ 0.6 - \frac{1}{15} \right] = 1.0889$$

$$\log_{10} \bar{S}^2 = \log_{10} \left( \frac{135.9799}{15} \right) = \log_{10}(9.0653) = 0.95738$$

$$B = \frac{2.3026}{1.0889} [(15)(0.95738) - 14.2504] = 0.2332$$

والقيمة 0.2332 تتوزع تبعا لتوزيع  $\chi^2$  بدرجات حرية 2، وهى غير معنوية حيث إن قيمة  $\chi^2$  الجدولية (جدول ٦ ملحق أ) عند درجات حرية 2 ومستوى معنوية 5% تساوى 5.99 وبالتالي لا يمكن رفض فرض العدم، أى أن التباينات متجانسة طبقاً لاختبار بارنلت.

### ١٢-٣-٢ اختبار $F_{\max}$ (Hartley, 1950)

وهذا اختبار بسيط فى إجراءاته، ويمكن إجراؤه باستخدام جدول ١٠ ملحق أ حيث إن:

$$F_{\max} = \frac{\text{largest variance of } k \text{ treatments}}{\text{lowest variance of } k \text{ treatments}} = \frac{\sigma^2 \text{ largest}}{\sigma^2 \text{ lowest}}$$

ويكشف عن قيمة  $F_{\max}$  عند مستوى معنوية وليكن  $\alpha$  ، k عدد المعاملات ودرجات حرية  $(n-1)$  حيث n: عدد الملاحظات فى كل معاملة (جدول ١٠ ملحق أ)، فإذا زادت المحسوبة عن الجدولية أو تساوت القيمتان أدى ذلك إلى رفض فرض العدم وتكون التباينات غير متجانسة.

فى مثال ١٢-١

$$F_{\max} = \frac{11.1867}{7.1227} = 1.57$$

وهى أقل من القيمة الجدولية 10.8 عند عدد معاملات 3 ودرجات الحرية 5 ومستوى معنوية 5% وعليه فلا يرفض فرض العدم ويستنتج من ذلك تجانس التباينات

وهي نفس النتيجة التي سبق الحصول عليها باختبار Bartlett ، ولكن الأخير أكثر كفاية من الأول. وإذا اختلف عدد المشاهدات في كل معاملة فيمكن استخدام  $n$  للمعاملة الأكثر عدداً، وهذا سيؤدي إلى رفض فرض العدم أكبر مما يجب، أي يجعل استنتاج المجرّب أكثر تحفظاً.

### ١٢-٣-٣ اختبار كوكران (1941) Cochran

وفيه تحسب

$$C = \frac{\sigma^2 \text{ largest}}{\sum_{j=1}^k \sigma_j^2}$$

حيث  $\sigma^2 \text{ largest}$  هو أعلى تقدير للتباين بين عدد  $k$  من المعاملات وأن  $\sum_{j=1}^k \sigma_j^2$  هو مجموع التباينات كلها وتقارن قيمة  $C$  المحسوبة بالقيمة المستخرجة من الجدول عند  $k$ ،  $(n-1)$ . ولقيمة  $C$  جداول خاصة للكشف عليها (جدول ١١ ملحق أ). وفي مثال ١٢-١ فإن

$$C = \frac{11.1867}{8.8867 + 7.1227 + 11.1867} = 0.4113$$

حيث  $k = 3$ ،  $(n-1) = 5$ ، وبالكشف في جدول ١١ ملحق أ يظهر أن القيمة 0.4113 أقل من القيمة الجدولية 0.7071 وعليه لا يرفض فرض العدم ويستتبط من ذلك أن التباينات متجانسة.

وجدير بالذكر أن اختبار  $F$ ،  $t$  لهما درجة كبيرة من الصمود والصلاحية حتى لو كان هناك تجاوز معقول عن تجانس التباينات، وأن الاختبارات الثلاثة المذكورة أعلاه حساسة لأي انحراف عن فرض تجانس التباينات لذا فإنه ليس من الإجراءات الروتينية أن يختبر المجرّب لتجانس التباينات إلا إذا كان هناك انحراف واضح عن التجانس أو أسباب أخرى تدعو المجرّب إلى هذا.

أما إذا أوضحت الاختبارات انحرافاً واضحاً عن فرض التجانس فقد يكون هذا مدعاة لأن يحاول المجرّب تحويل transformation قياسات التجربة، وهذا الموضوع سيناقش في نهاية هذا الباب.

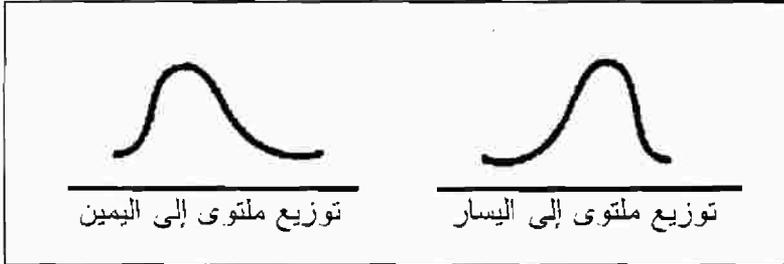
## ١٢-٤ طبيعية التوزيع Normally distributed

ويقصد أن الأخطاء  $e_{ij}$ 's موزعة توزيعاً طبيعياً. ويلاحظ أن الحيدة البسيطة عن التوزيع الطبيعي لا يكون لها تأثير كبير على اختبار  $F$ ،  $t$  ولكن يكون الأثر كبيراً إذا كان الالتواء skewness حاداً، أي أن معظم المشاهدات نحو يمين المنحنى أو نحو يساره. أو لو كان التفرطح kurtosis شديداً. ويمكن التعرف على هذا برسم البيانات كما سبق في الأبواب السابقة.

كما يمكن التوقع بأن التوزيع لن يكون طبيعياً من طبيعة البيانات نفسها فمثلاً البيانات العددية وليست القياسية غالباً مالا تتوزع حسب التوزيع الطبيعي وكذلك البيانات التي تؤخذ على هيئة نسبة مئوية تبعد قيمها عن 50% كثيراً ... الخ وسيتم مناقشة هذا عند الحديث عن التحويلات transformations.

### ١٢-٤-١ اختيار الالتواء Skewness

يكون منحنى التوزيع الطبيعي متماثلاً حول منتصفه. ولكن إذا تركزت المشاهدات أكثر إلى يمين المنحنى يطلق عليه ملتوى إلى اليسار وإذا كانت إلى اليسار يطلق عليه ملتوى إلى اليمين (شكل ١٢-١).



شكل ١٢-١ التوزيع الملتوى إلى اليسار والتوزيع الملتوى إلى اليمين

ويقال أن التوزيع التكراري موجب الالتواء إذا كان المتوسط أكبر من المنوال، ويقال أن التوزيع سالب الالتواء إذا كان المتوسط أقل من المنوال.

ويقدر معامل الالتواء بالمعادلة التالية:

$$\text{skew} = \frac{\text{mean} - \text{mode}}{\text{standard deviation}}$$

وحيث إنه أحيانا يصعب تقدير الـ mode وإن

$$\text{mean} - \text{mode} = 3(\text{mean} - \text{median})$$

فإن

$$\text{skew} = \frac{3(\text{mean} - \text{mode})}{\text{standard deviation}}$$

والمثال التالي يبين كيفية حساب واختبار الالتواء باستخدام العزوم حول المتوسط.

مثال ١٢-٢

جدول ١-١٢ يمثل الدرجة التي حصل عليها الطلبة في إحدى المواد الدراسية عدد 186 طالب، هل توزيع هذه الصفة في هذه العينة يتوزع حسب التوزيع الطبيعي؟

جدول ١-١٢ درجات الطلبة وتكراراتها في إحدى المواد الدراسية

$u^4$	$u^3$	$u^2$	$u$	التكرار $f$	الحد الأدنى للقسم
625	125	25	5	6	5
2401	343	49	7	4	7
6561	729	81	9	7	9
14641	1331	121	11	9	11
28561	2197	169	13	21	13
50625	3375	225	15	23	15
83521	4913	289	17	23	17
130321	6859	361	19	33	19
194481	9261	441	21	19	21
279841	12167	529	23	15	23
390625	15625	625	25	17	25
531441	19683	729	27	5	27
707281	24389	841	29	2	29
923521	29791	961	31	2	31

اختبار الالتواء:

$$h_1 = \frac{\sum fu}{n} = 17.634 \quad \sum fu = 3280$$

$$h_2 = \frac{\sum fu^2}{n} = 341.258 \quad \sum fu^2 = 63474$$

$$h_3 = \frac{\sum fu^3}{n} = 7060.086$$

$$\sum fu^3 = 1313176$$

$$h_4 = \frac{\sum fu^4}{n} = 153919.194$$

$$\sum fu^4 = 28628970$$

$$m_2 = \frac{\sum (X - \bar{X})^2}{n} = h_2 - h_1^2 = 30.30$$

$$m_3 = \frac{\sum (X - \bar{x})^3}{n} = h_3 - 3h_1h_2 + 2h_1^3 = -26.2795$$

$$\sqrt{b_1} = m_3 / m_2 \sqrt{m_2} = -26.2795 / 30.30 \sqrt{30.30} = -0.1576$$

وفى العينات التى تتبع التوزيع الطبيعى فإن الكمية  $\sqrt{b_1}$  تتوزع حسب التوزيع الطبيعى بمتوسط صفر وانحراف معيارى قدره  $\sqrt{6/n}$  أى أن الانحراف المعيارى هنا  $\sqrt{6/186} = 0.18$ .

والقيمة المقدرة  $\sqrt{b_1}$  هى (-0.158) وهى أقل من الانحراف المعيارى وعليه لا يرفض الفرض: أن هذه العينة اللتواء فيها منعدم. وتدل إشارة  $b$  على اتجاه اللتواء سالبا كان أم موجبا، فالسالبا يدل على تراكم المشاهدات نحو القيم المرتفعة من التوزيع، والموجب يدل على تراكم المشاهدات نحو القيم المنخفضة.

وجدير بالذكر أن ما ذكر من أن الكمية  $\sqrt{b_1}$  تتوزع طبيعياً بمتوسط صفر وانحراف معيارى  $\sqrt{6/n}$ ، وهذا الافتراض قريب من الصحة عندما تكون  $n$  قدرها 150 أو أكثر. وللعينات التى يتراوح عددها بين 25 إلى 200 فإن جدول ١٢ ملحق أ يعطى مستويات المعنوية 1%، 5% فى جهة واحدة one tail. فى المثال السابق فإن  $b_1 = -0.158$  أقل من تلك التى عند 1%، 5%. ولعينة مكونة من 200 فرد فإن القيمة الحرجة عند 1%، 5% هى 0.280، 0.403 على التوالى، وعليه لا يرفض فرض العدم بأن التوزيع طبيعى وهو مطابق لنفس الاستنتاج السابق.

ويتبين من جدول ١٢-١ أن الأرقام قد تصل إلى كميات كبيرة جداً خاصة تحت عمودى  $u^3$ ،  $u^4$  ويمكن اختزال قيمة  $u$  إما بطرح ثابت أو القسمة على ثابت أو الاثنى معاً وذلك لأن الكمية  $\sqrt{b_1}$  وكذلك مقياس التفرطح (كما سيأتى فيما بعد) مستقلان عن وحدة القياس وأى اختزال لا يؤثر فى حساباتهما.

### ١٢-٤-٢ اختبار التفطح Test of kurtosis

كما قيس الالتواء بواسطة متوسط القيمة  $(y - \mu)^3$  فى العشيرة وهى المسماة بالعزم الثالث، فإن التفطح يقاس بالعزم الرابع  $(y - \mu)^4$  مقسوماً على مربع التباين  $\sigma^4$ . وإذا كانت العينة تتبع حقاً التوزيع الطبيعي فإن القيمة المتوقعة لهذه النسبة = 3، فإذا كانت أكبر من 3 دل هذا على أن الشكل الناقوسى مديب القمة أكبر مما يجب وإن قل عن 3 دل على أن الناقوس مستوى القمة أكثر مما يجب.

وبالعودة إلى مثال ١٢-٢ وجدول ١٢-١ يمكن حساب ما يلى:

$$m_4 = h_4 - 4h_1h_3 + 6h_1^2h_2 - 3h_1^4 = 2545.7576$$

$$b_2 = m_4 / m_2^2 = 2.77288$$

$$g_2 = b_2 - 3 = 2.77288 - 3 = -0.2271$$

تتوزع الكمية  $g_2$  تقريباً حسب التوزيع الطبيعي بمتوسط قدره صفر وانحراف معيارى قدره  $\sqrt{24/n} = 0.359$  أى  $\sqrt{24/186}$ . الانحراف الذى قدره (-0.2271) لا يختلف معنوياً عن الصفر عند مستوى 5% وعليه يقبل الفرض القائل أن  $g_2 = 0$ ، أى أن التفطح هو ذلك المفروض أن يكون فى حالة التوزيع الطبيعي. وكون قيمة انحراف التفطح سالبة (أى أقل من 3) فإن ذلك يدل على أن منحني العينة مستوى القمة قليلاً ولكن ليس أكثر مما يجب تحت فرض 5% فى حالة التوزيع الطبيعي.

ويعتبر إجراء الاختبار على  $g_2$  مجاز حيث إنه يتطلب أن تكون  $n$  أكبر من 1000 حتى يقترب توزيع  $g_2$  من التوزيع الطبيعي. وجدول ١٣ ملحق أعطى قيمة مقربة أدق لكل من مستويات المعنوية 1%، 5% عندما يكون حجم العينة بين 200، 1000، وحيث إن توزيع  $g_2$  ملتوى فإن التوزيع معطى لكل من جهتي التوزيع مستقلاً. ففى المثال السابق حيث كانت  $b_2 = 2.77288$  وحجم العينة 186 فإن  $b_2$  أعلى من القيمة الحرجة الدنيا فى الجدول (القيمة الحرجة عند حجم العينة 200 هى 2.37، 2.51 عند 1%، 5% على التوالي)، وعليه لا يرفض فرض التوزيع الطبيعي. وهو نفس الاستنتاج السابق.

أما إذا قل عدد العينة عن 200 فإن Geary (1936) استنبط جداول بها القيم الحرجة للاختبارات حتى عدد العينة 11 وهى توافق إلى حد كبير قيم  $g_2$  المشروحة سابقاً (انظر أيضاً Snedecor and Chochran, 1987).

حل مثال ١٢-٢ باستخدام برنامج SAS لإجراء اختبار الالتواء والتفرطح

```
DATA GRADES;
INPUT GRADE REP @@;
CARDS;
5 6 7 4 9 7 11 9 13 21 15 23 17 23 19 33
21 19 23 15 25 17 27 5 29 2 31 2
PROC MEANS N MEAN STDERR SKEWNESS KURTOSIS T PRT;
VAR GRADE;
FREQ REP;
RUN;
* --- USING UNIVARIATE PROCEDURE ----;
PROC UNIVARIATE;
VAR GRADE;
FREQ REP;
RUN;
```

لاحظ

يمكن إجراء اختبار الالتواء والتفرطح إما باستخدام اختيار PROC MEANS أو اختيار .PROC UNIVARIATE.

نتائج التحليل

#### The MEANS Procedure

Analysis Variable : GRADE

N	Mean	Std Error	<u>Skewness</u>	<u>Kurtosis</u>	t Value	Pr >  t
186	17.6344086	0.4046066	-0.1568986	-0.2075433	43.58	<.0001

#### The UNIVARIATE Procedure

Variable: GRADE

Freq: REP

Moments

N	186	Sum Weights	186
Mean	17.6344086	Sum Observations	3280
Std Deviation	5.51809788	Variance	30.4494042
Skewness	-0.1568986	Kurtosis	-0.2075433
Uncorrected SS	63474	Corrected SS	5633.13978
Coeff Variation	31.2916526	Std Error Mean	0.40460657

١٢-٥ التجميعية Additivity

وهذه الخاصية أو الشرط قد يأخذ أكثر من معنى. ففي حالة تحليل التباين ذي الاتجاهين بدون تكرار المشاهدة كما في القطاعات العشوائية مثلاً وإذا كانت المعاملات ثابتة (نموذج ١) فإنه يجب فرض غياب التداخل بين المعاملة والقطاعات حتى يتسنى اختبارات الفروض الخاصة بالمعاملة باستخدام الخطأ من تحليل التباين. معنى ذلك أن أثر المعاملة يجمع على أثر القطاع ليحدثا أثرهما على المشاهدة. ولكن إذا ضرب الاثنان مثلاً فتكون خاصية التجميعية قد فقدت. وإذا وجد التداخل فعلاً وتم تجاهله ولم تجرى الاختبارات اللازمة لمعرفة مدى أهميته، فإن اختبار الفروض الخاصة بالمعاملات يصبح غير كفاء واختباراً ضعيفاً. ولاختبار وجود مثل هذا التداخل في حالة عدم وجود تكرار للمشاهدة داخل كل معاملة وكل قطاع سيتم مناقشته عند مناقشة تصميم القطاعات العشوائية. أما إذا تكررت المشاهدة فإنه يمكن اختبار خاصية التجميعية هذه لأنه سوف يكون هناك نوعان من الخطأ أحدهما يطلق عليه خطأ تجريبي experimental error وهو في حقيقته التداخل بين القطاعات والمعاملات وآخر يطلق عليه خطأ عيني sampling error وهو مقياس للتباين بين مشاهدات في نفس القطاع ونفس المعاملة. ونتاج قسمة متوسط مربعات الخطأ التجريبي على متوسط مربعات الخطأ العيني يعطى F لاختبار فرض عدم وجود تداخل بين القطاعات والمعاملات، وهذا أيضاً سيأتي شرحه تفصيلاً عند تقديم تصميم القطاعات العشوائية.

وفي التجارب الحيوية أو الزراعية يمكن أن ينتج التداخل من كثير من العلاقات بين المعاملات وطريقة تأثيرها على الكائن موضوع التجربة مثل ظاهرتا التعاضد synergism أو التعارض interference ويوجد أيضاً في الوراثة كما هو معروف بالتفوق بين الجينات epistasis والمثال الموضح في جدول ١٢-٢ يبين هذا التداخل.

ويتضح من جدول ١٢-٢ أن المشاهدة  $2 = a_1b_1$  عندما يكون سلوك المعاملات تجمعي وهي حاصل جمع  $a_1 + b_1$  ... وهكذا لبقية المشاهدات. وفي هذه الحالة فإن الفرق بين أي عمودين ثابت في أي صف وكذلك الفرق بين أي صفين ثابت في كل عمود.

بينما إذا كان السلوك ضربياً فإن المشاهدة  $a_3b_2$  مثلاً  $15 = (3)(5)$ ، ويكون الفرق بين أى صفتين غير متساو في الأعمدة، وكذلك الفرق بين الأعمدة في الصفوف المختلفة. وإذا علم الباحث أن مثل هذا السلوك سائد بين بيانات التجربة فيمكن علاج الأمر وتحوير البيانات، وذلك بأخذ لوغاريتمات القيم، الأمر الذى يوفر خاصية التجمعية كما هو واضح من الجدول السابق.

جدول ١٢-٢ مثال للتداخل بين العوامل وتوافر خاصية التجمعية

ملاحظات	العامل A			العامل B	
	$a_3 = 3$	$a_2 = 2$	$a_1 = 1$		
آثار تجمعية	4.00	3.0	2.0	تجمعى	$b_1 = 1$
آثار مضروبة	3.00	2.0	1.0	ضربى	
تحوير لوغاريتمى	0.48	0.3	0.0	لوغاريتم الضربى	
آثار تجمعية	8.00	7.0	6.0	تجمعى	$b_2 = 5$
آثار مضروبة	15.00	10.0	5.0	ضربى	
تحوير لوغاريتمى	1.18	1.0	0.7	لوغاريتم الضربى	

وفى حالة تحليل التباين ذى الاتجاه الواحد والذى يفترض فيه النموذج

$$Y_{ij} = \mu + t_i + e_{ij}$$

حيث  $\mu$  هي المتوسط و  $t_i$  أثر المعاملة و  $e_{ij}$  الخطأ، يمكن أن يحدث ألا تتوافر خاصية التجمعية كأن يكون النموذج الممثل للواقع مثلاً هو:

$$Y_{ij} = \mu t_i e_{ij}$$

أى مكونات النموذج مضروبة فى بعضها بدل أن تكون مجموعة على بعضها، وينتج عن هذا أيضاً عدم تجانس التباين.

ويمكن تلافى هذا الفرق للائسراطات اللازمة لتحليل التباين، وذلك بإجراء التحوير المناسب كما هو مبين فى جدول ١٢-٣.

جدول ١٢-٣ مثال لنموذج في اتجاه واحد مضروب العوامل والتحويل المناسب بفرض أن  $\mu = 1$

تحويل لوغاريتمي			المشاهدات الفعلية			
$t_3$	$t_2$	$t_1$	$t_3 = 6$	$t_2 = 4$	$t_1 = 2$	
0.7782	0.6021	0.3010	6	4	2	$e_{i1} = 1$
1.0792	0.9031	0.6021	12	8	4	$e_{i2} = 2$
1.2553	1.0792	0.7782	18	12	6	$e_{i3} = 3$
1.3802	1.2041	0.9031	24	16	8	$e_{i4} = 4$
0.0682	0.0682	0.0682	60	26.67	6.67	$S^2$

وواضح من جدول ١٢-٣ أنه إذا حللت النتائج بفرض نموذج تجميعي فإن هذا سوف يؤدي إلى تباينات غير متجانسة داخل المعاملات تراوحت قيمتها من 6.67 إلى (6)، بينما إذا أمكن التعرف على أن النموذج هو ضربى وعليه إجراء التحويل المناسب الذى تنتج عنه تباينات متجانسة تماماً، هذا بالطبع لأن المثال توضيحي بحث. وجدير بالذكر أن كثيراً من اشتراطات تحليل التباين مرتبطة ببعضها بمعنى أنه فى المثال السابق وجد أن البيانات الأصلية بياناتها مختلفة، وأن الخطأ مرتبط بقيمة المشاهدة، وأن خاصية التجمعية غير متوفرة، وبعمل التحويل المناسب تم علاج هذه النواحي الثلاث.

### ١٣-٦ تحويل البيانات Data transformations

المقصود بالتحويل هو إجراء نفس التغيير على كل البيانات الأولية بحيث تغير بعض خصائص البيانات بينما تسبقي الخصائص الأخرى كما هى. ويجرى التحويل عند القيام بتحليل التباين للأغراض الرئيسية التالية:

- ١- توفير تجانس التباينات؛
- ٢- توفير طبيعية التوزيع؛
- ٣- توفير خاصية التجمعية.

وكما سبق ذكره أن هناك عدة اشتراطات يجب توافرها حتى يمكن اختبار الفروض، وفى حالة عدم توافر بعض أو كل هذه الفروض فإنه ينصح أحياناً بإجراء تحويل فى البيانات أملاً فى التوصل إلى شكل محدد من البيانات يتوافر فيه الاشتراطات المطلوبة ولكن يظل محتفظاً بخصائص البيانات من فروق بين المعاملات

والاتجاهات الأخرى المطلوب دراستها. ولأن اختبار F أقل تأثراً بعدم توافر شرطي طبيعية التوزيع وتجانس البيانات، فإن إجراء التحويل للغرض الثالث هو أدعى من إجرائه للغرضين الأولين. والغرض الثالث هذا هام جداً توافره خصوصاً في بعض التصميمات الإحصائية التي قد لا يتوافر فيها تكرار المشاهدة مثل تصميم القطاعات العشوائية randomized block.

وكما سبق القول، فإن التحويل ممكن أن يعالج أكثر من نقص في آن واحد، كما في جدول ١٢-٣، وإن كان من الممكن إجراء التحويل المناسب لمعالجة بعض النواقص كما في حالة ارتباط التباين بمتوسط المعاملة، إلا أنه ليست القاعدة أن يجد المجرّب التحويل المناسب في كل الحالات. ويذكر (Kirk, 1995) أنه لا يوجد تحويل مناسب إذا وجدت أي من الحالات التالية:

- ١- متوسط المعاملات متساوية تقريباً ولكن التباينات غير متجانسة.
- ٢- متوسطات المعاملات تختلف بطريقة مستقلة عن التباينات.
- ٣- تجانس التباينات ولكن شكل توزيع متوسطات المعاملات غير متجانس.

ونظرياً إذا علم السلوك الرياضي للملاحظات تماماً، فإنه يمكن إيجاد التحويل المناسب ولكن قلما يتوافر مثل هذا الوضع في الحياة العملية. وهناك عدة طرق لمحاولة إيجاد التحويل المناسب ولكن سيتم سرد التحويلات التي وجدت مناسبة في حالات معينة. وفي حالة عدم إمكان التوصل إلى التحويل المناسب يمكن إتباع طرق تحليلية غير معتمدة على توزيع معين وهي ما يطلق عليها الإحصاءات اللامعلمية nonparametric statistics أو الإحصاءات حرة التوزيع distribution-free statistics وهي تخصص أبعد من مجال هذا المؤلف.

### ١٢-٦-١ تحويل الجذر التربيعي Square root transformation

يتناسب في بعض أنواع البيانات التباين مع متوسط المعاملة كما هو الحال في توزيع بواسون حيث  $\sigma^2 = \mu$ . وينشأ مثل هذا التوزيع عندما يكون المتغير التابع (أو المشاهدة) ناتج عن عد أفراد أو أشياء أو أحداث والتي احتمال حدوثها بسيط مثل عدد بويضات طفيل معين في أمعاء حيوان معالج ضد الطفيليات أو عدد الحيوانات المنوية الشاذة في سائل منوى محفوظ أو عدد خلايا معينة في الدم ... الخ. في مثل هذه الحالات يحور كل رقم إلى الجذر التربيعي له، وإن كانت هناك أصفار فيكون التحويل باستخدام  $Y = \sqrt{Y + 0.5}$  ثم يجرى التحليل واختبارات الفروض على البيانات المحورة. وترد متوسطات المعاملات وانحرافات القياسية فقط إلى وحداتها الأصلية بإحداث تحويل عكسي.

وجداول ١٢-٤ يبين هذا التحويل لعدد الحيوانات المنوية الشاذة في مساحة معينة من الشريحة تحت الميكروسكوب لثلاثة مخففات.

جدول ١٢-٤ تأثير ثلاثة مخففات على عدد من الحيوانات المنوية الشاذة في مساحة محددة من الشريحة

البيانات المحورة $y = \sqrt{y + 0.5}$			البيانات الأصلية $y$			
المخفف			المخفف			
الثالث	الثاني	الأول	الثالث	الثاني	الأول	
4.85	3.54	2.55	23	12	6	
3.54	2.92	0.71	12	8	0	
3.54	2.12	2.92	12	4	8	
4.53	2.92	2.12	20	8	4	
3.54	3.81	2.12	12	14	4	
<b>4.00</b>	<b>3.06</b>	<b>2.084</b>	<b>15.8</b>	<b>9.2</b>	<b>4.4</b>	المتوسط
<b>0.4095</b>	<b>0.4265</b>	<b>0.7016</b>	<b>28.2</b>	<b>15.20</b>	<b>8.80</b>	$S^2$

المتوسطات محسوبة من البيانات المحورة ثم مردودة إلى وحداتها الأصلية، أي يربع المتوسط ثم يطرح منه 0.5، هي 3.84، 8.86، 15.5 على التوالي).

وواضح من هذا المثال أن أعلى تباين كان أعلى أكثر من ثلاث مرات أي  $28.2/8.8 = 3.2$  قدر أقل تباين قبل التحويل بينما أصبح أعلى تباين أقل من ضعف أقل تباين أي  $0.7016/0.4095 = 1.7$ .

### ١٢-٦-٢ التحويل اللوغاريتمي Logarithmic transformation

يستخدم هذا التحويل عندما يكون هناك تناسب بين الانحرافات المعيارية وامتوسطات، ويجرى التحويل كما يلي:

$$y' = \log_{10} Y$$

أو  $y' = \log_{10}(Y + 1)$  عندما يكون هناك قيم قدرها صفراً أو صغيرة جداً.

ويفيد هذا التحويل إذا ما كانت معاملة معينة مثلاً تزيد عن معاملة أخرى بنسبة ثابتة، وليس بقدر ثابت فمثلاً في جدول ١٢-٢ القيمة  $a_2$  دائماً ضعف القيمة  $a_1$  (الفرق بينهما 1 تحت  $b_1$ ، 5 تحت  $b_2$ ) وأن  $a_3$  ثلاثة أضعاف  $a_1$  و 50% أعلى من  $a_2$  (الفرق 2، 1 تحت  $b_1$  على التوالي، 10، 5 تحت  $b_2$  على التوالي). وذلك ينطبق تحت  $b_1, b_2$ . لذا فإن التحويل اللوغاريتمي حول الزيادة بالنسبة إلى زيادة بالإضافة فأصبح الفرق بين  $a_2$  و  $a_1$  هو 1 في كل الحالات ... وهكذا. وكذلك في جدول ١٢-٣ فالمشاهدات الفعلية في  $t_3$  كانت دائماً ثلاث مرات قدر  $t_1$ ، 50% أعلى من  $t_2$  ولكن بعد التحويل اللوغاريتمي أصبح هذا الفرق 0.311، وهكذا في جميع الأحوال.

ويستخدم التحويل اللوغاريتمي أيضاً عندما يكون المتغير التابع هو زمن الاستجابة وتكون البيانات ملتوية إلى يمين التوزيع. مثل فترة الاستجابة reaction time وهي الفترة ما بين إدخال الذكر إلى الإناث والتلقيح.

### ١٢-٦-٣ التحويل بالمقلوب Inverse transformation

ويستخدم في مثل حالات التحويل اللوغاريتمي عندما يكون هناك تناسب بين متوسطات المعاملات وانحرافات المعيارية، ويكون التحويل  $y' = 1/Y$ . وعندما تكون إحدى قيم  $Y$  مساوية للصفر فإن  $y' = 1/(Y+1)$ .

### ١٢-٦-٤ تحويل مقلوب جيب الزاوية Arcsin or angular transformation

يفيد هذا التحويل عندما يكون المتوسط والتباين متناسبين وأن التوزيع يتبع "دو الحدين". وتتحقق مثل هذه الظروف عندما يكون العدد ثابتاً مثلاً والمتغير هو عدد مرات النجاح من العدد الكلي. كأن يعطى للذكر عدد ثابت من الإناث ثم يسجل عدد الإناث المخصبة للمقارنة بين الذكور المختلفة. كذلك يفيد هذا التحويل عندما تكون البيانات مسجلة على هيئة نسب مئوية وخصوصاً إذا كانت أقل من 30% أو أعلى من 70% مثل النسبة المئوية للرطوبة في اللحم أو نسبة الدهن في الذبيحة ... الخ. أما نسبة التصافي في الذبيحة مثلاً فهي لا تحتاج عادة إلى تحويل، لأن النسبة المئوية لها لا تتخطى الحدين 30، 70% ويجرى هذا التحويل عن طريق  $y' = 2 \arcsin \sqrt{y}$  حيث  $y$  معبر عنها كنسبة. ويعطى جدول ١٤ ملحق أ القيم المحورة من 0.001 وحتى 0.999 ويقترح أن تعطى  $1/2n$  أو  $1/4n$  في حالة  $y$  تساوى صفر،  $(1-1/2n)$  في حالة  $y = 1$ ، حيث  $n$  هي عدد المشاهدات عندما تكون  $n$  أقل من 50.

مثال ١٢-٤

البيانات التالية تمثل نسبة مئوية للأنسجة الدهنية إلى الوزن الكلي لإحدى القطعيات في ذبيحة الأغنام تحت ثلاثة أنظمة غذائية مختلفة.

جدول ١٢-٥ نسبة الأنسجة الدهنية في إحدى القطعيات بذبيحة الأغنام في ثلاثة معاملات

البيانات المحورة			البيانات الأصلية			
معاملة ١	معاملة ٢	معاملة ٣	معاملة ١	معاملة ٢	معاملة ٣	
0.6435	0.7670	1.1152	0.10	0.14	0.28	
0.7075	0.6435	1.2239	0.12	0.10	0.33	
0.6094	0.8763	1.2661	0.09	0.18	0.35	
0.6761	0.8230	1.1152	0.11	0.16	0.28	
0.6761	0.7954	1.0004	0.11	0.15	0.23	
0.7954	0.7075	1.0004	0.15	0.12	0.23	
<b>0.6847</b>	<b>0.7688</b>	<b>1.202</b>	<b>0.113</b>	<b>0.142</b>	<b>0.283</b>	المتوسط
<b>0.0041</b>	<b>0.0069</b>	<b>0.0122</b>	<b>0.0004</b>	<b>0.0008</b>	<b>0.0025</b>	S <sup>2</sup>

ويبين من جدول ١٢-٥ أن  $F_{max} = 0.0025/0.0004 = 6.25$  وهى معنوية. يتدل على عدم تجانس التباينات.

بينما بعد تحويل البيانات  $F_{max} = 0.0122/0.0041 = 3.0$  وهى غير معنوية.

مثال ١٢-٥

حل مثال ١٢-٤ باستخدام برنامج SAS

```
DATA TRANS;
INPUT TRT FAT @@;
TFAT = 2*ARSIN(SQRT(FAT));
CARDS;
1 0.1 1 0.12 1 0.11 1 0.11 1 0.15
2 0.14 2 0.1 2 0.18 2 0.16 2 0.15 2 0.12
3 0.28 3 0.33 3 0.35 3 0.28 3 0.23 3 0.23
PROC MEANS MEAN VAR;
```

VAR FAT TFAT;  
BY TRT;  
RUN;

نتائج التحليل

The MEANS Procedure

----- trt=1 -----

Variable	Mean	Variance
FAT	0.1180000	0.000370000
TFAT	0.6997288	0.0033720

----- trt=2 -----

Variable	Mean	Variance
FAT	0.1416667	0.000816667
TFAT	0.7687848	0.0069335

----- trt=3 -----

Variable	Mean	Variance
FAT	0.2833333	0.0024667
TFAT	1.1201828	0.0121622

-----

## تمارين الباب الثاني عشر

١٢-١ في إحدى التجارب لقياس عدد بيض الديدان الاسطوانية في أمعاء أربع سلالات من الأغنام وجدت النتائج التالية علماً بأن كل سلالة ممثلة بعدد عشر تقديرات للبيض.

السلالة	متوسط عدد البيض	الانحراف المعياري S
١	2250	210
٢	12100	4051
٣	8650	2025
٤	9160	2516

- أ- افحص العلاقة بين المتوسط والانحراف المعياري  
ب- اختبر ما إذا كانت التباينات متجانسة

١٢-٢ البيانات التالية لطول فترة الحمل في الجاموس المصري، افحص ما إذا كانت هذه العينة تتبع التوزيع الطبيعي باستخدام اختبار الالتواء والتفرطح.

التكرار	الحد الأدنى للقسم (يوم)
10	296
18	301
71	306
131	311
156	316
99	321
24	326
6	331
<b>515</b>	<b>إجمالي</b>

١٢-٣ التالي هو عدد البكتيريا في ١ سم<sup>٢</sup> من اللبن لثلاث بقرات في ثلاث فترات مختلفة

رقم البقرة	عند الحلب	بعد ٢٤ ساعة من الحلب	بعد ٤٨ ساعة من الحلب
١	12000	14000	57000
٢	13000	20000	65000
٣	21500	31000	106000

١- احسب المتوسط والتباين لكل من الثلاث فترات وافحص العلاقة بين التباين والمتوسط. حور البيانات لوغاريتمياً ثم أعد حساب التباين والمتوسط وقارن بين الحالتين، ثم ناقش النتائج.

٢- حلل التباين على كل من البيانات الأصلية وتلك المحورة وقارن النتائج.