١٢ الافتراضات الخاصة بتحليل التباين

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 فى البابين التاسع والحادى عشر وكما سيتضح فيما بعد فى الباب الخامس عشر أنه لإجراء تحليل التباين وتكوين الجدول الخاص به فإن الافتراض الوحيد المطلوب هو أن تكون درجات الحرية بكل مصدر مستقلة خطياً أى 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. والتى عادة ما يستخدم فيها جداول F ، t أو χ^2 ... الخ. فلإجراء اختبار هذه الفروض باستخدام تلك الجداول لابد أن يستوثق المجرب أن هناك عدة شروط تتوافر فى المادة التجريبية. وأهم هذه الشروط هى: عشوائية المعاينة، استقلال المشاهدات، تجانس التباينات، طبيعية التوزيع، والتجمعية. وسوف يتم فى 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.

١٢-١ عشوائية المعاينة Random sampling

هذا شرط أساسى لصحة اختبار أيٍ فرض. فالعينة التى سوف يجرى عليها اختبار الفرض لابد أن تكون مأخوذة عشوائياً من عشيرة محددة التعريف والذى سوف يردُّ الاستنباط عليها. فإذا أراد المجرب 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 المجرب، 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 المجرب، افترض أن المجرب 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 بالتدريج 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، فالفروق بين المعاملات ربما ستحتوى أيضاً على فرق فى الحملان نفسها وهذا طبعاً لم يقصده المجرب. حتى وإن كان ذلك "السيناريو" المذكور مبالغ فيه ولكن لن يكلف المجرب جهداً كبيراً إذا هو أجرى التعشية بالطرق سالفة الذكر.

أيضاً كثيراً ما يتم وضع الحيوانات المولودة، فى محطة تجارب مثلاً، فى قوائم حسب تاريخ ميلادها مثلاً، فإذا أخذ المجرب المجموعة الأولى للمعاملة الأولى والثانية للثانية ... وهكذا، فإنه يدخل عامل العمر بطريقة غير مقصودة، وهذا أيضاً قد يؤدى إلى تحيز نتائج التجربة. وهكذا يمكن سرد عديد من الأمثل المتوقعة وغير المتوقعة التى قد تؤدى إلى تحيز النتائج دون قصد من المجرب. وصمام الأمان فى هذا هو إجراء التعشية إجراءاً سليماً، وليس للتعشية العامة اختبار معين، ولكن يمكن اختبار مثلاً إذا كانت المعاملات موزعة عشوائياً على عامل معين مثلاً كالوزن أو العمر ... الخ. وعدم توافر هذا الشرط يؤدى إلى حالة يكون المجرب غير واثق فيها من إرجاع الظواهر إلى مسبباتها الحقيقية والتى تحيز النتائج.

١٢-٢ استقلالية المشاهدات Independence of observations

المقصود هنا باستقلالية المشاهدات هو استقلال الخطأ فى هذه المشاهدات أى e_{ij} فى النموذج الإحصائى 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 المجرب لديه أربعة حيوانات فى كل معاملة وكل معاملة وضعها فى حظيرة، وفى كل معاملة يوضع كمية ثابتة من العليقة، والمطلوب اختبار الفرق بين العلائق. ففى هذه الحالة يمكن أن ينشأ وضع هو أنه إذا استهلكت بعض الحيوانات فى 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، وفى ذلك حيدة عن الاستقلال وستكون هذه الحيدة عن استقلال الـ e_{ij} بصورة أكبر إذا ما كانت مساحات أماكن الغذاء (المداود) محدودة الأمر الذى سيؤدى إلى تصارع (تناطح) الحيوانات والأقوى هو الذى سيفوز بنصيب أكبر من العليقة على حساب الأضعف وهو ما يطلق عليه bullying effect. ولهذا يجب أن يفطن المجرب لمثل هذه الأمور مسبقاً، وأن يحاول علاجها مسبقاً طبقاً لإمكانياته كأن يعطى لكل حيوان كمية من العليقة مستقلاً عن بقية الحيوانات إذا سمح المكان بهذا أو يكثر من المساحات المخصصة للتغذية "الطوايل" أو تقدم العليقة مجزأة على عدة مرات إذا سمحت أغراض التجربة بذلك أو قص قرون الحيوانات حتى تقلل من هذا التأثير.

وهناك نوع من ترابط الأخطاء دأب كثير من باحثى الإنتاج الحيوانى 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 فى الحسبان عند التخطيط أو التحليل يؤدى إلى ارتباط بين الأخطاء. فالحيوان الذى يقاس نفسه عدة مرات فى الشتاء ثم فى الربيع ثم فى الصيف ثم فى الخريف، إذا كان تنفسه

سريعاً بالنسبة لبقية الحيوانات فى أحد هذه المواسم، يحتمل أن يكون سريعاً نسبياً أيضاً فى بقية المواسم الآخرى وفى هذا ارتباط بين الأخطاء. ومن هذا أيضاً ظاهرة المعاومة فى الأشجار والتى ينشأ عنها ارتباط سالب ... الخ، وسيعالج هذا الموقف فى الباب الخامس عشر. وقد ينشأ الارتباط أيضاً فى حالة عدم التعشية السليمة. ويمكن أخذ فكرة عن هذا الارتباط من حساب e_{ij} لكل مشاهدة فهذه القيم يجب ألا تأخذ اتجاهاً معيناً فى 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 رتبت المشاهدات عشوائياً أو تكون منتظمة التكرار ويتوالى السالب منها والموجب فى نظام معين. وهناك اختبارات للكشف عن استقلالية الأخطاء منها اختبار ديربن واطسون Dubrin/Watson statistic. وليس هناك من أسلوب لمعالجة البيانات حتى يتوافر فيها شرط الاستقلال هذا، ولكن الذى يجب عمله هو التعشية السليمة مع اختيار التصميم الإحصائى 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 فى النموذج الإحصائى، ففى الأبواب السادس والتاسع والعاشر فإن تجانس التباين كان شرط مسبق لإجراء اختبارات الفروض. فإذا كان الهدف هو اختبار الفرق بين معاملتين مثلاً طبقاً للنموذج $Y_{ij} = \mu + a_i + e_{ij}$ فالفرض هنا يعنى أنه إذا قدر التباين داخل المعاملة الأولى بقيمة $\hat{\sigma}^2_{e1}$ وداخل المعاملة الثانية بقيمة $\hat{\sigma}^2_{e2}$ فإن هذين التقديرين هما تقدير لقيمة مشتركة واحدة هى $\hat{\sigma}^2_e$ أى أن الأخطاء فى المعاملة الأولى وتلك التى فى المعاملة الثانية يتبعان (أو مأخوذتان من) نفس العشيرة.

وهناك عدة طرق لاختبار فرض تجانس التباينات أى

$$H_o : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma_e^2$$

١٢-٣-١ اختبار بارتلت لتجانس التباين

Bartlett's (1937) test for homogeneity of variance

فى هذا الاختبار تقدر قيمة B كما يلى:

$$B = \frac{2.3026}{C}\{[\sum(n_i - 1)]\log_{10} \overline{S}^2 - \sum(n_i - 1)\log_{10} S_i^2\} \qquad (١٢-١)$$

حيث

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

S_i^2 : تقدير غير متحيز لتباين (σ_i^2) المعاملة i

$\overline{S}^2$: متوسط التباين

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

n_i : عدد المشاهدات فى المعاملة i

K : عدد التباينات

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

مثال ١٢-١

إذا كانت أوزان الحملان فى ثلاث مجموعات كالتالى:

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 أى:

$$H_\circ : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$$

يمكن إجراء اختبار بارتلت لتجانس التباين كالتالى:

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

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

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

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

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

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

١٢-٣-٢ اختبار F_{max} (Hartley, 1950)

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

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

ويكشف عن قيمة F_{max} عند مستوى معنوية وليكن α ، 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

وهى نفس النتيجة التى سبق الحصول عليها باختبار Bartlett ، ولكن الأخير أكثر كفاية من الأول. وإذا اختلف عدد المشاهدات فى كل معاملة فيمكن استخدام n للمعاملة الأكثر عدداً، وهذا سيؤدى إلى رفض فرض العدم أكبر مما يجب، أى يجعل استنتاج المجرب أكثر تحفظاً.

١٢-٣-٣ اختبار كوكران Chochran (1941)

وفيه تحسب

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

حيث σ^2 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). وفى مثال ١٢-١ فإن

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

حيث k = 3، 5 = $(n-1)$، وبالكشف فى جدول ١١ ملحق أ يظهر أن القيمة 0.4113 أقل من القيمة الجدولية 0.7071 وعليه لا يرفض فرض العدم ويستنبط 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 لأى انحراف عن فرض تجانس التباينات لذا فإنه ليس من الإجراءات الروتينية أن يختبر المجرب لتجانس التباينات إلا إذا كان هناك انحراف واضح عن التجانس أو أسباب أخرى تدعو المجرب إلى هذا.

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

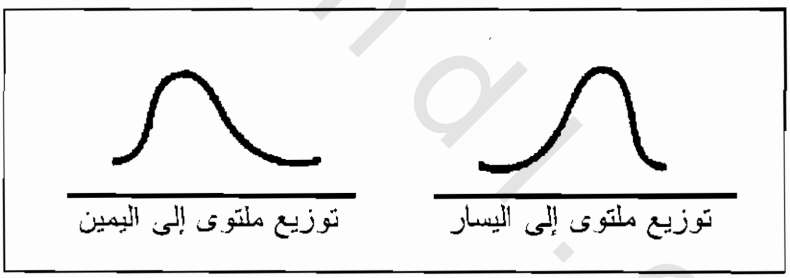
١٢-٤ طبيعية التوزيع Normally distributed

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

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

١٢-٤-١ اختبار 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 التكرارى موجب الالتواء إذا كان المتوسط أكبر من المنوال، ويقال أن التوزيع سالب الالتواء إذا كان المتوسط أقل من المنوال.

ويقدر معامل الالتواء بالمعادلة التالية:

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

وحيث إنه أحيانا يصعب تقدير الـــ mode وإن

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

فإن

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

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

مثال ١٢-٢

جدول ١٢-١ يمثل الدرجة التى حصل عليها الطلبة فى احدى المواد الدراسية عدد 186 طالب، هل توزيع هذه الصفة فى هذه العينة يتوزع حسب التوزيع الطبيعى؟

جدول ١٢-١ درجات الطلبة وتكراراتها فى إحدى المواد الدراسية

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

اختبار الالتواء:

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

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

$$h_3 = \frac{\Sigma fu^3}{n} = 7060.086 \qquad \sum fu^3 = 1313176$$

$$h_4 = \frac{\Sigma fu^4}{n} = 153919.194 \qquad \sum fu^4 = 28628970$$

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

$$m_3 = \frac{\Sigma (X - \overline{x})^3}{n} = h_3 - 3h_1 h_2 + 2h_1^3 = -26.2795$$

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

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 الالتواء فيها منعدم. وتدل إشارة b على اتجاه الالتواء سالباً كان أم موجباً، فالسالب 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. ففى 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 الاثنين 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 التفرطح Test of kurtosis

كما قيس الالتواء بواسطة متوسط القيمة $(y-\mu)^3$ في العشيرة وهي المسماة بالعزم الثالث، فإن التفرطح يقاس بالعزم الرابع $(y-\mu)^4$ مقسوماً على مربع التباين σ^4. وإذا كانت العينة تتبع حقاً التوزيع الطبيعي فإن القمة المتوقعة لهذه النسبة = 3، فإذا كانت أكبر من 3 دل هذا على أن الشكل الناقوسي مدبب 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 يلي:

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

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

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

تتوزع الكمية g_2 تقريباً حسب التوزيع الطبيعي بمتوسط قدره صفر وانحراف معياري قدره $\sqrt{24/n}$ أي $\sqrt{24/186} = 0.359$. الانحراف الذي قدره (0.2271 -) لا يختلف معنوياً عن الصفر عند مستوى 5% وعليه يقبل الفرض القائل أن $g_2 = 0$، أي أن التفرطح هو ذلك المفروض أن يكون في حالة التوزيع الطبيعي. وكون قيمة انحراف التفرطح سالبة (أي أقل من 3) فإن ذلك يدل على أن منحنى العينة مستوى القمة قليلاً ولكن ليس أكثر مما يجب تحت فرض 5% في 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 أ يعطي قيماً 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. ففي المثال السابق حيث كانت $b_2 = 2.77288$ وحجم العينة 186 فإن b_2 أعلى من القيمة الحرجة الدنيا في الجدول (القيمة الحرجة عند حجم العينة 200 هي 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 وهي توافق إلى حد كبير قيم g_2 المشروحة سابقاً (انظر أيضاً Snedecor and Chochran, 1987).

مثال ١٢-٣

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	Skewness	Kurtosis	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 التجمعية Additivity

وهذه الخاصية أو الشرط قد يأخذ أكثر من معنى. ففى حالة تحليل التباين ذى الاتجاهين بدون تكرار المشاهدة كما فى 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 التجمعية قد فقدت. وإذا وجد التداخل فعلاً وتم تجاهله ولم تجرى الاختبارات اللازمة لمعرفة مدى أهميته، فإن اختبار الفروض الخاصة بالمعاملات يصبح غير كفء واختباراً ضعيفاً. ولاختبار وجود مثل هذا التداخل فى 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 التجمعية هذه لأنه سوف يكون هناك نوعان من الخطأ أحدهما يطلق عليه خطأ تجريبى experimental error وهو فى حقيقته التداخل بين القطاعات والمعاملات وآخر يطلق عليه خطأ عينى sampling error وهو مقياس للتباين بين مشاهدات فى نفس القطاع ونفس المعاملة. وناتج قسمة متوسط مربعات الخطأ التجريبى على متوسط مربعات الخطأ العينى يعطى F لاختبار فرض عدم وجود تداخل بين القطاعات والمعاملات، وهذا أيضاً سيأتى شرحه تفصيلاً عند تقديم تصميم القطاعات العشوائية.

وفى 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 فى الوراثة كما هو معروف بالتفوق بين الجينات epistasis والمثال الموضح فى جدول ١٢-٢ يبين هذا التداخل.

ويتضح من جدول ١٢-٢ أن المشاهدة $a_1b_1 = 2$ عندما يكون سلوك المعاملات تجمعى وهى حاصل جمع $a_1 + b_1$ ... وهكذا لبقية المشاهدات. وفى هذه الحالة فإن الفرق بين أى عمودين ثابت فى أى صف وكذلك الفرق بين أى صفين ثابت فى كل عمود.

بينما إذا كان السلوك ضربياً فإن المشاهدة a_3b_2 مثلاً $15 = (5)(3)$، ويكون الفرق بين أى صفين غير متساو فى الأعمدة، وكذلك الفرق بين الأعمدة فى 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

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

وفى حالة تحليل التباين ذى الاتجاه الواحد والذى يفترض فيه النموذج

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

حيث μ هى 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 للاشتراطات اللازمة لتحليل التباين، وذلك بإجراء التحوير المناسب كما هو مبين فى جدول ١٢-٣.

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

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

وواضح من جدول ١٢-٣ أنه إذا حللت النتائج بفرض نموذج تجميعى فإن هذا سوف يؤدى إلى تباينات غير متجانسة داخل المعاملات تراوحت قيمتها من 6.67 إلى 60، بينما إذا أمكن التعرف على أن النموذج هو ضربى وعليه إجراء التحوير المناسب الذى تنتج عنه تباينات متجانسة تماماً، هذا بالطبع لأن المثال توضيحى بحت. 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 التحوير المناسب تم علاج هذه النواحى الثلاث.

١٢-٦ تحوير البيانات Data transformations

المقصود بالتحوير هو إجراء نفس التغيير على كل البيانات الأولية بحيث تغير بعض خصائص البيانات بينما تستبقى الخصائص الأخرى كما هى. ويجرى التحوير 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 تحوير فى البيانات أملا 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 شرطى طبيعية التوزيع وتجانس البيانات، فإن إجراء التحوير للغرض الثالث هو أدعى من إجرائه للغرضين الأولين. والغرض الثالث هذا هام جداً توافره خصوصاً فى بعض التصميمات الإحصائية التى قد لا يتوافر فيها تكرار المشاهدة مثل تصميم القطاعات العشوائية randomized block.

وكما سبق القول، فإن التحوير ممكن أن يعالج أكثر من نقص فى آن واحد، كما فى جدول ١٢-٣، وإن كان من الممكن إجراء التحوير المناسب لمعالجة بعض النواقص كما فى حالة ارتباط التباين بمتوسط المعاملة، إلا أنه ليست القاعدة أن يجد المجرب التحوير المناسب فى كل الحالات. ويذكر (Kirk, 1995) أنه لا يوجد تحوير مناسب إذا وجدت أى 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 الرياضى للمشاهدات تماماً، فإنه يمكن إيجاد التحوير المناسب ولكن قلما يتوافر مثل هذا الوضع فى الحياة العملية. وهناك عدة طرق لمحاولة إيجاد التحوير المناسب ولكن سيتم سرد التحويرات التى وجدت مناسبة فى حالات معينة. وفى حالة عدم إمكان التوصل إلى التحوير المناسب يمكن إتباع طرق تحليلية غير معتمدة على توزيع معين وهى ما يطلق عليها الإحصاءات اللامعلمية nonparametric statistics أو الإحصاءات حرة التوزيع distribution-free statistics وهى تخصص أبعد من مجال هذا المؤلف.

١٢-٦-١ تحوير الجذر التربيعى Square root transformation

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

وجدول ١٢-٤ يبين هذا التحوير لعدد الحيوانات المنوية الشاذة فى مساحة معينة من الشريحة تحت الميكروسكوب لثلاثة مخففات.

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

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

المتوسطات محسوبة من البيانات المحورة ثم مردودة إلى وحداتها الأصلية، أى يربع المتوسط ثم يطرح منه 0.5، هى 3.84، 8.86، 15.5 على التوالى).

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

١٢-٦-٢ التحوير اللوغاريتمى Logarithmic transformation

يستخدم هذا التحوير عندما يكون هناك تناسب بين الانحرافات المعيارية والمتوسطات، ويجرى التحوير كما يلى:

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

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

ويفيد هذا التحوير إذا ما كانت معاملة معينة مثلاً تزيد عن معاملة أخرى بنسبة ثابتة، وليس بقدر ثابت فمثلاً فى 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 و50% أعلى من a_2 (الفرق 2، 1 تحت b_1 على التوالى ، 10، 5 تحت b_2 على التوالى). وذلك ينطبق تحت b1, b2. لذا فإن التحوير اللوغاريتمى حول الزيادة بالنسبة إلى زيادة بالإضافة فأصبح الفرق بين a_2 و a_1 هو 1 فى كل الحالات ... وهكذا. وكذلك فى جدول ١٢-٣ فالمشاهدات الفعلية فى t_3 كانت دائماً ثلاث مرات قدر t_1، 50% أعلى من t_2 ولكن بعد التحوير اللوغاريتمى أصبح هذا الفرق 0.311، وهكذا فى جميع الأحوال.

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

١٢-٦-٣ التحوير بالمقلوب Inverse transformation

ويستخدم فى مثل حالات التحوير اللوغاريتمى عندما يكون هناك تناسب بين متوسطات المعاملات وانحرافاتها المعيارية، ويكون التحوير $y' = 1/Y$. وعندما تكون إحدى قيم Y مساوية للصفر فإن $y' = 1/(Y+1)$.

١٢-٦-٤ تحوير مقلوب جيب الزاوية Arcsin or angular transformation

يفيد هذا التحوير عندما يكون المتوسط والتباين متناسبين وأن التوزيع يتبع "ذو الحدين". وتتحقق مثل هذه الظروف عندما يكون العدد ثابتاً مثلاً والمتغير هو عدد مرات النجاح من العدد الكلى. كأن يعطى للذكر عدد ثابت من الإناث ثم يسجل عدد الإناث المخصبة للمقارنة بين الذكور المختلفة. كذلك يفيد هذا التحوير عندما تكون البيانات مسجلة على هيئة نسب مئوية وخصوصاً إذا كانت أقل من 30% أو أعلى من 70% مثل النسبة المئوية للرطوبة فى اللحم أو نسبة الدهن فى الذبيحة ... الخ. أما نسبة التصافى فى الذبيحة مثلاً فهى لا تحتاج عادة إلى تحوير، لأن النسبة المئوية لها لا تتخطى الحدين 30, 70% ويجرى هذا التحوير 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 $1/4n$ فى حالة y تساوى صفر، $(1-1/2n)$ فى حالة y = 1، حيث n هى عدد المشاهدات عندما تكون n أقل من 50.

مثال ١٢-٤

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

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

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

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

بينما بعد تحوير 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 الثانى عشر

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

السلالة	متوسط عدد البيض	الانحراف المعيارى 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 فى الجاموس المصرى، افحص ما إذا كانت هذه العينة تتبع التوزيع الطبيعى باستخدام اختبار الالتواء والتفرطح.

الحد الأدنى للقسم (يوم)	التكرار
296	10
301	18
306	71
311	131
316	156
321	99
326	24
331	6
إجمالى	**515**

١٢-٣ التالى هو عدد البكتيريا فى ١ سم^{٣} من اللبن لثلاث بقرات فى 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.