3-2: التحويلات Transformations :

إن عدم توفر بعض الفروض السابقة في البيانات التي يراد تحليل التباين لها يؤدي إلى قرارات خاطئة ، لذلك فأن هناك طرق مختلفة لمعالجة ذلك منها استخدام التحويلات المناسبة وهذا يتطلب معرفة التوزيع الفعلي للبيانات حتى نختار طريقة التحويل المناسبة ثم نجرى بعد ذلك تحليل التباين على البيانات الجديدة المحولة ، وفيما يلي أهم طرق التحويل :

1-3-2: التحويل أللوغاريتمي : Logarithmic Transformation :

يستخدم هذا التحويل عندما تكون هناك علاقة نسبية بين الانحرافات المعيارية للعينات (المجموعات) وبين متوسطاتها (أي أن معامل الاختلاف ثابت $\frac{S}{\bar{y}}$) وكذلك عندما نلاحظ أن التأثيرات الأساسية ليست تجميعية بل نسبية أو تضاعفية . وفي حالة وجود قيم تساوي الصفر فيكون التحويل بإضافة 1 للبيانات .

مثال (2-3) :

البيانات التالية تمثل إحدى التجارب المقامة لمعرفة فيما إذا كانت هناك فروق معنوية بين المعالجات الأربعة a , b , c , d التي استخدمت في التجربة.

	d	c	b	a
	4000	43300	1520	895
	8600	32800	1610	540
	8260	28800	1900	1020
	9830	34600	1350	470
	7600	27800	980	428
	9650	32800	1710	620
	8900	28100	1930	760
	6060	18900	1960	537
	10200	31400	1840	845
	15500	39500	2410	1050
	9250	29000	1520	387
	7900	22300	1685	497
المجموع	112750	369300	20415	8049
المتوسط	9395.83	30775	1701.25	970.75
الانحراف المعياري	2254.99	6462.75	340.25	228.05

يلاحظ أن هناك علاقة طردية بين المتوسط والانحراف المعياري للمعالجات لذا يمكن استخدام التحويل أللوغاريتمي للبيانات وكما يلي :

لتحويل القيمة الأولى ضمن المعالجة A والتي هي (895) نجد أن قيمة اللوغاريتم لها هي 2.951823 فإذا كتبت هذه القيمة بمرتبتين عشريتين فقط بعد الفارزة وكذلك لبقية القيم نجد أن قيم المفردات لجميع المعالجات المحولة إلى اللوغاريتم كما في الجدول الآتي :

d	c	b	a
4.04	4.64	3.18	2.95
3.93	4.52	3.21	2.73
3.92	4.46	3.28	3.01
3.99	4.54	3.13	2.67
3.88	4.44	2.99	2.63
3.98	4.52	3.23	2.79
3.95	4.45	3.29	2.88
3.78	4.28	3.29	2.73
4.01	4.50	3.16	2.93
4.19	4.60	3.38	3.02
3.97	4.46	3.18	2.59
3.90	4.35	3.23	2.70

ثم يجري تحليل التباين على هذه البيانات المحولة ..

2-3-2 :التحويل باستخدام الجذر ألتربيعي * Square Root Transformation

يستخدم هذا التحويل عندما تكون البيانات عبارة عن أعداد تشير إلى حصول بعض الحوادث النادرة (ذات الاحتمال الصغير) مثل عدد بقع التعفن على أوراق النباتات أو عدد البكتريا على قطعة معدنية أو عدد الوحدات المعابة في إنتاج معين . وعليه فأن هذه البيانات تقترب من توزيع بواسون (Poisson) ، كذلك بالإمكان استخدام هذا التحويل عندما يكون تباين كل معالجة يتناسب مع المتوسط أي أن $S^2 = C\bar{y}$.

حيث تكون C مساوية إلى الواحد عندما يكون التوزيع هو توزيع بواسون وغالبا تكون C أكبر من الواحد . هذا وعندما تكون عدد القيم أقل من (10) نستخدم التحويل $y+\frac{1}{2}$ بدلا من y .

* يمكن الاستفادة من الجدول 9 في الملحق الذي يتضمن تحويل القيم الى الجذر التربيعي.

مثال (4-2) :

البيانات التالية عبارة عن عدد النباتات الحمراء في الشوفان لتجربة استخدمت فيها خمسة معالجات هي a , b , c , d , e لتقليل هذه النباتات ، حيث وزعت هذه المعالجات على وحدات التجربة بصورة عشوائية .

المعالجات

	a	b	c	d	e
	438	538	77	17	18
	442	422	61	31	26
	319	377	157	87	77
	380	315	52	16	20
المتوسط	394.75	413	86.75	37.75	35.25
التباين	3352.9	8868.7	2300.25	1124.9	786.25

يلاحظ في الجدول أعلاه أن هناك اختلاف في التباين مما يقودنا إلى الشك بأن التباينات غير متجانسة وإذا ما حسب متوسط مربعات الخطأ فأنه سيكون كبير عند اختبار الفرق بين المعالجات e , d , c وصغيرا بالنسبة إلى b , a وعليه بعد أخذ الجذر ألتربيعي لبيانات الجدول أعلاه فأن هذه الاختلافات ستقل ، وكما موضح في الجدول أدناه .

(على سبيل المثال الجذر ألتربيعي لقيمة المشاهدة الأولى للمعالجة (a) هو $\sqrt{438} = 20.92844$ وهذه النتيجة تقرب إلى مرتبة عشرية واحدة بعد الفارزة وكذلك بالنسبة لبقية القيم) .

المعالجات

	a	b	c	d	e
	20.9	23.2	8.8	4.1	4.2
	21.0	20.5	7.8	5.6	5.1
	17.9	19.4	12.5	9.3	8.8
	19.5	19.7	7.2	4.0	4.5
المتوسط	19.8	20.2	9.1	5.8	5.6
التباين	2.1	5.3	5.6	6.1	4.55

وعليه يمكن أجراء تحليل التباين على البيانات المحولة.

2-3-3: التحويل الزاوي (التحويل باستخدام النسب المثلثية) :

Arc sins Transformation

يستخدم هذا التحويل عندما تتبع البيانات توزيع ذي الحدين (Binomial) وعادة تكون البيانات على شكل نسب مئوية ، حيث تحول بعد ذلك الى دالة جيب الزاوية العكسية ويمكن ان نتوصل الى ذلك باستخدام العلاقات المعروفة التالية :

P : يمثل احتمال النجاح

q : يمثل أحتمل الفشل

$q = 1-p$

وبما أن $p+q = 1$

كما ان $\sin^2 h + \cos^2 h = 1$

لتكن $P=\sin^2 h$

فعليه $h = Sin^{-1}\sqrt{p}$

ومن جدول النسب المثلثية أو جداول خاصة نجد جيب الزاوية العكسية .

مثال (5-2) :

الجدول الآتي يبين نتائج تجربة لمكافحة الحشرات التي تصيب الذرة استخدمت فيها أربع معالجات (a,b,c,d) وزعت عشوائيا على وحدات التجربة والبيانات في الجدول تبين النسبة غير المتضررة من الحاصل .

المعالجات

a	b	c	d
42.4	33.3	8.5	16.6
34.3	33.3	21.9	19.3
24.1	5.0	6.2	16.6
39.5	26.3	16.0	2.1
55.5	30.2	13.5	11.1
49.1	28.6	15.4	11.1

وباستخدام التحويل الزاوي حسب العلاقة $h = Sin^{-1}\sqrt{p}$ تحول إلى القيم كما في الجدول الآتي : (كمثال فإن النسبة a=42.4 يتم تحويلها كالآتي $Sin^{-1}\sqrt{0.424} = 40.6$)

a	b	c	d
40.6	35.2	17	24
35.8	35.2	27.9	26.1
29.4	12.9	14.4	24
38.9	30.9	23.6	8.3
48.2	33.3	21.6	19.5
44.5	32.3	23.1	19.5

4-2: تحليل التباين لمعيار واحد: One-Way Classification Analysis of Variance

لغرض توضيح فكرة تجزئة التباين إلى مركبتين لنفترض أن لدينا (t) من المجموعات (**المعالجات**، العينات، الطرق، الأصناف، ...) وفي كل مجموعة (n_i) من المفردات (المشاهدات) فأن قيم المشاهدات أو الاستجابات يمكن أن يرمز لها كما في الجدول (1-2) الآتي :

جدول رقم (2-1) يبين رموز قيم(الاستجابات) المشاهدات في كل مجموعة

المتوسط $\bar{y}_{i.}$	المجموع $y_{i.}$	المشاهدات او المفردات 1 2 … j … n_i	المجموعات
$\bar{y}_1$	$y_{1.}$	y_{11} y_{12} … y_{1j} … y_{1n1}	1
$\bar{y}_2$	$y_{2.}$	y_{21} y_{22} … y_{2j} … y_{2n2}	2
.	.	.	.
.	.	.	.
$\bar{y}_{i.}$	$y_{i.}$	y_{i1} y_{i2} … y_{ij} … y_{ini}	i
.	.	.	.
$\bar{y}_{t.}$	$y_{t.}$	y_{t1} y_{t2} … y_{tj} … y_{tnt}	t
	$y_{..}$		

حيث أن :

y_{ij} : تمثل قيمة أو نتيجة المفردة (المشاهدة) ضمن المجموعة (i) والتي ترتيبها (j) .

$y_{i.}$: مجموع نتائج جميع المفردات (مشاهدات) المجموعة (i) ويعبر عنه رياضياً

$$y_{i.} = \sum_{j=1}^{ni} y_{ij}$$

$\bar{y}_{i.}$: المتوسط لنتائج المجموعة (i) ويعبر عنه رياضيا .

$$\bar{y}_{i.} = \frac{y_{i.}}{n_i} = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$$

$y_{..}$: المجموع العام لنتائج جميع المفردات ويعبر عنه رياضيا .

$$y_{..} = \sum_{i=1}^{t} y_{i.} = \sum_{i}^{t} \sum_{j}^{n_i} y_{ij}$$

$\bar{y}_{..}$: المتوسط (الوسط الحسابي) العام لجميع نتائج لمفردات ويعبر عنه رياضيا :

$$\bar{y}_{..} = \frac{y_{..}}{N = \sum_{i=1}^{t} n_i} = \frac{\sum y_{i.}}{N} = \frac{\sum_i \sum_j y_{ij}}{N}$$

وهنا يجب أن نفترض أن المشاهدات (y_{ij}) تتوزع طبيعيا بمتوسطات هي:

($\mu_1, \mu_2, \ldots, \mu_t$) للمجموعات $(1,2,..,t)$ على التوالي وأن الانحراف المعياري لهذه المجموعات نفترضه متساوي وثابت ونرمز له بالرمز (σ).

والسؤال الذي يهمنا هنا هو هل هناك فروقات معنوية (جوهرية) بين المجموعات على ضوء المشاهدات في كل مجموعة ؟ أن فرضية العدم التي تقابل هذا التساؤل هي :

$$Ho : \mu_1 = \mu_2 = \ldots = \mu_t$$

من الممكن اختبار فرضية العدم أعلاه بأخذ متوسطين في كل مرة وتطبيق اختبار (t) كما تعلمنا سابقا ألا أننا سنقوم بالاختبار بالشكل التالي :

<u>أولا:</u> يمكن اعتبار المجموعات على أنها t من المجتمعات التي لها نفس التباين وهو σ^2 وإذا افترضنا أن ($u_1 = u_2 = \ldots = u_t$) فأن هذه المجتمعات ربما تعتبر مجتمع واحد كبير (حيث u هو المتوسط العام) وأن المجموعات t هي عينات مأخوذة من هذا المجتمع الكبير . ومن معلوماتنا السابقة فأن التقدير غير المتحيز لتباين المجتمع يمكن استخراجه من خلال تجميع تباينات عدد العينات المأخوذة منه ، وفي الحالة أعلاه فأن تباينات العينات هي :

$$S_1^2 = \frac{1}{n_1 - 1} \sum_j^{n_1} (y_{1j} - \bar{y}_{1.})^2$$

$$S_2^2 = \frac{1}{n_2 - 1} \sum_j^{n_2} (y_{2j} - \bar{y}_{2.})^2$$

.

$$S_t^2 = \frac{1}{n_t - 1} \sum_j^{n_t} (y_{t.} - \bar{y}_{t.})^2$$

وأن التباين المجمع (Pooled) هو :

$$\hat{\sigma}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 + \ldots + (n_t - 1)S_t^2}{(n_1 + n_2 + \ldots + n_t) - t} \quad \ldots\ldots\ldots(2-6)$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{t}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})^2}{N-t} \quad ....(2-7)$$

ثانيا : يمكن تقدير التباين $\hat{\sigma}^2$ باستخدام العلاقة التالية :

$$\sigma^2_{\bar{Y}} = \frac{\sigma^2}{n} \rightarrow \sigma^2 = n\sigma^2_{\bar{Y}}$$

وعليه فمن خلال تقدير $\sigma^2_{\bar{Y}}$ يمكن تقدير σ^2

للعينة الأولى : $\sigma^2 = n_1\sigma^2_{\bar{Y}_1} = n_1(\bar{Y}_{1.} - \bar{Y}...)^2$

للعينة الثانية : $\sigma^2 = n_2\sigma^2_{\bar{Y}_2} = n_2(\bar{Y}_2 - \bar{Y}...)^2$

للعينة t : $\sigma^2 = n_t\sigma^2_{\bar{Y}_t} = n_t(\bar{Y}_{t_t} - \bar{Y}...)^2$

وللتقدير غير المتحيز نستخدم درجات حرية (t-1) بدلا من (t) وعليه :

$$\boxed{\hat{\sigma}^2 = \frac{1}{t-1}\sum_{i-1}^{t} n(\bar{Y}_{i.} - \bar{Y}..)^2} \quad ......(2-8)$$

وهي الطريقة الثانية في تقدير σ^2

ملاحظة : في حالة تساوي عدد المشاهدات للعينات $(n_1 = n_2 = ...n_t = n)$

فأن التقدير يمكن ان يكون :

$$\sigma^2 = \frac{1}{t-1}\sum_{i-1}^{t} n(\bar{Y}_{i.} - \bar{Y}..)^2$$

وحيث أن كل من الصيغتين (2-7) و (2-8) يقدران تباين المجتمع فنتوقع ان تكون قيمة الصيغة (2-7) وقيمة الصيغة (2-8) لا تختلفان جدا أي أننا نتوقع ان تكون النسبة بين الصيغتين قريبة الى الواحد وكما مبين أدناه :

$$F = \frac{\frac{1}{t-1}\sum_{i-1}^{t} n(\bar{Y}_{i.} - \bar{Y}..)^2}{\frac{1}{N-t}\sum_{i=1}^{t}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_{i.})^2} \quad ....(2-9)$$

إلا أن قيمة (F) ستكون أكبر من واحد عندما تكون فرضية العدم غير صحيحة .

السؤال المطروح الآن هو إلى أي مدى يمكن أن تكون قيمة (F) أكبر من الواحد ليمكن عندها أن نقرر أن فرضية العدم يجب أن ترفض ؟ .

وللإجابة على هذا السؤال ينبغي معرفة توزيع (F) . وقبل ذلك سنتطرق إلى طريقة أخرى في تقدير تباين المجتمع .

ثالثا : إذا كانت فرضية العدم هي :

$$Ho: \mu_1 = \mu_2 = \dots = \mu_t$$

فإن هناك طريقة ثالثة في تقدير تباين المجتمع σ^2 لأنه في هذه الفرضية سنعتبر إن العينات (K) كعينة كبيرة حجمها هو : $n = n_1 + n_2 + \dots + n_t$

وعليه فأن التقدير غير المتحيز إلى σ^2 هو :

$$\sigma^2 = \frac{\sum_{i=1}^{t}\sum_{j=1}^{n}(y_{ij} - \bar{y}..)^2}{N-1} \quad ......(2-10)$$

ويتضح من خلال الصيغ (2-7) ، (2-8) ، (2-10) إن هناك علاقة في قيمة البسط في هذه الصيغ حيث أن :

$$\sum_i \sum_j (y_{ij} - \bar{y}..)^2 = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 + \sum_i \sum_j (\bar{y}_{i.} - \bar{y}...)^2 \quad ... (2-11)$$

حيث أن :

$\sum\sum(y_{ij} - \bar{y}..)^2$ يمثل مجموع مربعات الانحرافات الكلي .

$\sum\sum(y_{ij} - \bar{y}_{i.})^2$ يمثل مجموع مربعات الانحرافات داخل العينات (المجموعات)

$\sum\sum(\bar{y}_{i.} - \bar{y}..)^2$ يمثل مجموع مربعات الانحرافات بين العينات(المجموعات)

إن المعادلة (2-11) يشار لها على أنها المتطابقة الأساسية لتحليل التباين وهي تعبر عن الفكرة القائلة بأن مجموع المربعات الكلي لانحرافات قيم المشاهدات عن الوسط الحسابي العام يساوي مجموع مربعات انحرافات متوسطات العينات (المجموعات) عن الوسط الحسابي العام مضافا إليه مجموع مربعات انحرافات القيم داخل كل عينة (مجموعة) ، وأن

هذه المتطابقة تصح سواء كانت فرضية العدم ($H_o : \mu_1 = \mu_2 = \ldots = \mu_t$) بهذا الشكل أو غيره .

إن المعادلة (2-11) يمكن اشتقاقها بصورة رياضية كما يلي :

مجموع المربعات الكلي انحرافات قيم المشاهدات عن الوسط الحسابي العام هو : $\sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2$, الآن إذا أضفنا وطرحنا داخل القوس الكمية $\bar{y}_{i.}$ فتصبح الصيغة كالأتي :

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

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

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

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

وبما أن مجموع انحرافات القيم عن وسطها الحسابي الحقيقي يساوي صفر فالصيغة تصبح :

$$\underbrace{\sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2}_{total} = \underbrace{\sum_i \sum_j (Y_{ij} - \bar{Y}_{i.})^2}_{within} + \underbrace{\sum \sum (\bar{Y}_{i.} - \bar{Y}_{..})^2}_{between}$$

أو الكلي = داخل + بين

رابعا : تطبيق توزيع F

إذا كان هناك متغيران W ، Q كل منهما يتوزع بصورة مستقلة توزيع χ^2 (مربع كاي) وإذا قسم كل منهما على درجة الحرية المناظرة له فأن النسبة الناتجة (قسمة احدهما على الآخر) تسمى (نسبة التباين) وأن التوزيع لها هو توزيع F .وعند تطبيقنا لهذا المفهوم في اختبار تساوي تباينين ($\sigma_1^2 = \sigma_2^2$) نرى إن النسبة للتقدير غير المتحيز σ_1^2 , σ_2^2 تحقق الشرط الخاص بنسبة التباين وعليه يكون لها توزيع F .

وفي مثالنا السابق والذي نريد فيه اختبار تساوي الأوساط فأننا قد أوجدنا نسبة مجموع المربعات بين ومجموع المربعات داخل أوقسمة بين على داخل أو

$$F = \frac{\frac{1}{t-1} \sum_i n_i (\bar{y}_{i.} - \bar{y}_{..})^2}{\frac{1}{N-t} \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2} \qquad (2-12)$$

حيث أن البسط والمقام هما تقديران غير متحيزين لتباين المجتمع σ^2 وأن هذه النسبة أيضا لها توزيع F ويمكن تبسيط هذه النسبة كما يلي :

F= تقدير التباين (بين) / تقدير التباين (داخل)

وعندما تكون أوساط المجتمع (المجموعات) غير متساوية فأن تقدير التباين من مجموع مربعات (بين) مساويا إلى $\sigma^2 + C$ حيث $C > zero$ وبالنتيجة فأن نسبة التباين F ستكبر وبعبارة أخرى إذا كانت F كبيرة فأن هناك سبب يجعلنا نشك في تساوي الأوساط أو في كون العينات مأخوذة من نفس المجتمع .

أن العمليات الحسابية الحقيقية لمجاميع المربعات المؤشرة في المعادلة (2-11) تكون أكثر سهولة إذا بسطت أولا وأعيدت كتابتها بدلالة المجموع لكل مجموعة (من المجموعات t) وكالآتي :

مجموع المربعات الكلي سيكون :

$$\sum\sum (y_{ij}-\bar{y}_{..})^2 = \sum\sum y_{ij}^2 - \frac{(\sum\sum y_{ij})^2}{N} \qquad \ldots(2\text{-}13)$$

حيث أن الكمية $\frac{(\sum\sum y_{ij})^2}{N} = \frac{y_{..}^2}{N}$ يرمز لها في بعض الأحيان بالرمز (C.F) وهي تعني معامل التصحيح (Correction Factor) مجموع المربعات لبين العينات (المجموعات) سيكون :

$$\sum_i\sum_j (\bar{y}_{i.}-\bar{y}_{..})^2 = \sum_i \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N} \qquad \ldots(2\text{-}14)$$

مجموع المربعات لداخل العينات (المجموعات) يستخرج من طرح مجموع المربعات لبين المجموعات من مجموع المربعات الكلي أو وفق الصيغة التالية :

$$\sum\sum (y_{ij}-\bar{y}_{i.})^2 = \sum\left[\sum y_{ij}^2 - \frac{y_{i.}^2}{n_i}\right] \qquad \ldots(2-15)$$

أن المعلومات لمجاميع المربعات للمركبات أعلاه يمكن وضعها في الجدول(2-2) التالي الذي يسمى جدول تحليل التباين لمعيار واحد .

جدول (2-2) تحليل التباين لمعيار واحد .

مصادر التباين Sources of variation (S.O.V)	درجات الحرية Degrees of freedom (d.f)	مجموع المربعات Sum of squares (S.S)	متوسط المربعات Mean squares (M.S)	F
بين المجموعات (between groups)	t-1	$SSG=\sum_{i=1}^{t}\frac{y_{i.}^{2}}{n_i}-\frac{y_{..}^{2}}{N}$	$MSG=\frac{SSG}{t-1}$	$\frac{MSG}{MSw}$
داخل المجموعات أو الخطأ (within groups or Error)	N-t	$SSW=\sum\left[\sum y_{ij}^{2}-\frac{y_{i.}^{2}}{n_{i.}}\right]$	$MSw=\frac{SSw}{N-t}$	
Total الكلي	N-1	$\sum_i\sum_j y_{ij}^{2}-\frac{y_{..}^{2}}{N}$		

ويتم مقارنة قيمة F المحسوبة في الجدول أعلاه مع قيمة F الجدولية (الجدول3 في الملحق)، وحينما تكون المحسوبة اكبر من الجدولية فترفض H_0 .[أن مركبة داخل المجموعات عادة يطلق عليها الخطأ ومجموع مربعاتها كما قلنا يمكن استخراجه بصيغة مبسطة بالاعتماد على المتطابقة الأساسية في تحليل التباين لمعيار واحد وأن مجموع المربعات لداخل المجموعات في هذا الجدول يساوي مجموع المربعات الكلي مطروحا منه مجموع المربعات لبين المجموعات ، وما قيل عن مجـ المربعات يقال عن درجات الحرية ، أي أن درجة الحرية لمركبة داخل المجموعات تساوي درجة حرية الكلي مطروحا منها درجة حرية بين المجموعات] .

وأخيرا لابد من معرفة أن تسمية تحليل التباين لمعيار واحد يقصد بها أن التجربة تقام باتجاه واحد أي أن كل مشاهدة من المشاهدات الموجودة واقعة تحت تأثير مجموعة واحدة. وعند دراستنا للتصاميم لاحقا ستمثل كل مجموعة مستوى من مستويات العامل الذي ندرسه ويطلق عليها تسمية معالجة وعليه سيرجع سبب التسمية إلى أن هناك عامل واحد في التجربة هو موضوع البحث والدراسة . أما إذا كان هناك عاملين في التجربة فأن لكل عامل مجموعة من المعالجات وستكون كل مشاهدة واقعة تحت تأثير معالجتين (مجموعتين) ويسمى تحليل التباين لهذه التجارب بتحليل التباين لمعيارين كما سنأخذه لاحقا .

2-4-1: النماذج الخطية في تحليل التباين لمعيار واحد

2-4-1-1: النموذج المقيد (الثابت) أو النموذج I) : Fixed Model (Mode 1)

لنفرض أن هناك t من المكائن (للتبسيط لتكن t=3) وهي C , B , A التـي تنـتج منتج معين وزنه (y) غرام ونرغب في البحث فيما إذا كان هناك فرق في متوسط الــوزن للإنتاج الخاص بالمكائن الثلاثة وهنا يجب أن نفترض :

1 . أن أوزان المنتج الخاص بكل ماكنة يتوزع طبيعيا بمتوسط مقداره μ_i وتباين σ^2 على التوالي .

2 . تباينات المجتمعات (أي المكائن الثلاثة متساوية أي أنهـا جميعـا σ^2 وأن مـشكلتنا الإحصائية هي لاختبار فرضية العدم .

$$H_0: \mu_1 = \mu_2 = \mu_3$$

أما الفرضية البديلة فهي :

$$H_1: \mu_1 \neq \mu_2 \neq \mu_3$$

وبما أننا مهتمون في الفروقات بين الأوساط μ_i سنضع النموذج يوضح بصورة صـريحة هذه الفروق.

I. المتوسط العام يعرف كما يلي:

$$\mu = \frac{1}{t}\sum_{i=1}^{t} \mu_i = \frac{1}{3}\sum_{i=1}^{3} \mu_i$$

II. الانحرافات الخاصة بـــ μ_i عن μ تعرف كما يلي:

$$T_i = \mu_i - \mu \quad , i = 1, 2, 3,\ldots \quad (2.16)$$

أي أن تختلف عنه بمقدار (T_3, T_2, T_1) بعبارة أخرى :

$$\mu_1 = \mu + T_1$$
$$\mu_2 = \mu + T_2$$
$$\mu_3 = \mu + T_3$$

أن T_i تسمى التأثيرات أو الانحرافات نتيجة المعالجات وبالنسبة إلى مثالنا ، فأن T_i هو تأثير الماكينة i وأن هذه التأثيرات تجميعية حيث أن :

$$\sum_{i=1}^{t} T_i = \sum_{i=1}^{t} (\mu_i - \mu) = 0$$

هذه النتيجة هي بسبب الطريقة التي عرفنا بها المتوسط العـام μ . وهنـا يجـب ملاحظة انه عندما تكون ($T_i =$ صفر) فأن المتوسطات تكون مساوية للمتوسط العام أي أنه :

$\mu = \mu_1 = \mu_2 = \mu_3$.

ويعتبر هذا أسلوبا أخر في تفسير فرضية العدم ، كما يلاحظ هنا أيضا أن T_i هي مجموعة من الثوابت تمثل الانحرافات الخاصة بـ μ_i عن μ أي أن هناك t من قيم T_i .

الآن لنركز انتباهنا على ناتج الماكينة i فأن وزن المنتج y_{ij} يتوزع طبيعيا بمتوسـط مقداره μ_i وتباين σ^2 وبما أن هذه الأوزان تتوزع حول متوسط الوزن فربمـا سـيكون هناك انحرافات لقيم y_{ij} عن μ_i ناتجة الاختلافات العشوائية وهذه الاختلافات هي :

$e_{ij} = y_{ij} - \mu_i$ …… (2-17)

وتسمى هذه أيضا الأخطاء العشوائية ويفترض توزيعها طبيعيا بمتوسط مقداره صفر وتباين σ^2 وتكون هذه الأخطاء مستقلة . وعليه يمكن كتابة النموذج بالشكل التالي :

$y_{ij} = \mu_i + e_{ij}$ …. (2-18)

وبتعويض المعادلة (2-16) في (2-18) ينتج أن :

$y_{ij} = \mu + T_i + e_{ij}$ ……….(2-19)

وهذا هو النموذج المقيد لتحليل التباين لمعيار واحد .

2-4-1-2: النموذج العشوائي (نموذج II) Random Model(Model II) :

في النموذج المقيد كان اهتمامنا على المكائن الثلاثة فقط ولكن لنأخذ بنظر الاعتبار المشكلة التالية :

أفترض أن أحد المنتجين رغب في معرفة فيما إذا كانت المكائن المنتجة من هذا النوع لها نفس الخصائص . فعمليا لا يستطيع تجربة جميع المكائن وعليه سيختار عينة عشوائية ، على سبيل المثال بحجم(4) مكائن ومن كل ماكينة اخذ (5) مشاهدات وللتوضيح أكثر نفترض أن عدد المكائن من هذا النوع (10) مكائن . النقطة التي يجب ملاحظتها هنا أن المكائن الأربعة هي عينة عشوائية من مجتمع المكائن وسيكون اهتمامنا ليس الاختلاف بهذه المكائن الأربعة فقط ولكننا نرغب في الاستدلال عن الاختلافات في المكائن العشرة (المجتمع) وعليه فان نموذج التأثيرات العشوائية هو :

$$\mu_i = \mu + T_i$$
$$y_{ij} = \mu_i + e_{ij}$$
$$y_{ij} = \mu + T_i + e_{ij}\,, \quad i = 1,2,3,4$$
$$j = 1,2,3,4,5$$

وبما ان المكائن الأربعة عينة عشوائية من مجتمع المكائن فان μ_i تتوزع طبيعيا بمتوسط مقداره μ وتباين هو σ_A^2 وأن σ_A^2 يعرف كما يلي :

$$\sigma_A^2 = E(\mu_i - \mu)^2 = \frac{1}{t}\sum_{i=1}^{t}(\mu_i - \mu)^2$$

وعليه فأن σ_A^2 يوضح الاختلافات في μ_i حول μ وكلما كان σ_A^2 كبيرا كلما كانت الاختلافات في μ_i كبيرة وهذا يؤدي إلى أن الاختلافات بين μ_i نفسها كبيـرة ، وبعبـارة أخـرى عنـدما $\sigma_A^2 \rightarrow$ صفر وهـذا يعنـي أن $\mu_i \rightarrow \mu$ وعنـدما تكـون $\sigma_A^2 =$ صفر فأن $\mu_i = \mu$ وأنه ليس هناك اختلاف بين μ_i وعليه فأن الفرضية التي سنختبرها في هذه الحالة يمكن وضعها كما يلي :

$$H_o : \sigma_A^2 = \text{صفر}$$
$$H_1 : \sigma_A^2 \neq \text{صفر}$$

وبصورة عامة يمكن أن نلخص ما ذكر في أعلاه كالآتي:

1 . أن خطوات حساب فقرات تحليل التباين وبالتالي أجراء اختبار F هي واحدة لكـل مـن النموذج الثابت (المقيد) والنموذج العشوائي .

2 . أن الاختلاف بين النموذجين يمكن تأشيره في النقطتين الآتيتين :

1)) في حالة استخدام النموذج الثابت فأن الهدف من التجربة هو تقدير تأثير المعالجـــة (Ti) ومقارنة تأثير المعالجات وبالتالي مقارنة المتوسطات للمعالجات، أما في حالة استخدام النموذج العشوائي فأن الهدف هو تقدير مكون التباين σ_A^2 .

2)) عند ظهور أن قيمة F الحسابية في جدول تحليل التباين معنوية فأن الخطوة اللاحقة فـــي حالة التأثيرات الثابتة هي مقارنة متوسطات المعالجات، أما في حالة التأثيرات العشوائية فأن الخطوة اللاحقة يمكن أن تكون حساب قيمة تباين تأثير المعالجات.

كذلك يمكن توضيح المقارنة بين النموذجين وبصورة مبسطة وكالآتي :

وحدة المقارنة Comparison Unit	النموذج الثابت Fixed Model		النموذج العشوائي Random Model	
Assumptions الافتراضات	$\sum T_i = \sum_{i=1}^{t} (\mu_i - \mu) = 0$ تأثيرات المعالجات ، قيم ثابتة		Ti~ **NID** $(0, \sigma_T^2)$ تأثيرات المعالجـات متغيرات عشوائية تتوزع طبيعيا وبصورة مستقلة بوسط حسابي يساوي صفر وتباين يساوي σ_T^2	
Analysis التحليل	**The same steps**		**نفس الخطوات السابقة**	
مكونات التباين components E.M.S	**S.O.V**	**E.M.S**	**S.O.V**	**E.M.S**
	Treatments	$\sigma_e^2 + r\frac{\sum T_i^2}{t-1}$	Treatments	$\sigma_T^2 + r\,\sigma_T^2$
	Error	σ_e^2	Error	σ_e^2
الفرضيــــة المختبــــرة Hypothesis tested	Ho : Ti =0 (for all t's) Ho : Ti ≠ 0 (for all t's)		$H_o : \sigma_T^2 = 0$ $H_1 : \sigma_T^2 \neq 0$	

$$\sigma_A^2 = E(\mu_i - \mu)^2 = \frac{1}{t}\sum_{i=1}^{t} (\mu_i - \mu)^2$$

وعليه فأن σ_A^2 يوضح الاختلافات في μ_i حول μ وكلما كان σ_A^2 كبيرا كلما كانــت الاختلافات في μ_i كبيرة وهذا يؤدي إلى أن الاختلاف بين μ_i نفسها كبيرة ، وبعبارة أخرى عندما $\sigma_A^2 \rightarrow$ صفر هذا يعني أن $\mu_i \rightarrow \mu$ وعندما تكون $\sigma_A^2 =$ صفر ، فــأن $\mu_i = \mu$ وأنه ليس هناك اختلاف بين μ_i وعليه فأن الفرضية التي ســنختبرها فــي هــذه الحالة يمكن وضعها كما يلي :

$$H_o : \sigma_A^2 = \text{صفر}$$

$$H_1 : \sigma_A^2 \neq \text{صفر}$$

تجربة (2-1) :

اجري امتحان في مادة الإحصاء لمجموعات من طلبة المرحلة الأولى في أقسام الإحصاء في خمس جامعات(Universities) وكانت الدرجات التي حصلوا عليها كما في الجدول (2-3) أدناه:

جدول (2-3) يبين درجات الطلبة في امتحان مادة الإحصاء

الجامعات						$y_{i.}$
الموصل	57	57	65	71	69	319
صلاح الدين	62	77	62	68	68	337
بغداد	79	95	92	69	83	418
واسط	67	57	63	70	65	322
البصرة	69	79	89	85	50	372

$y_{..} = 1768$

المطلوب:عمل جدول تحليل التباين واختبار معنوية الفرق بين المجموعات (الجامعات). أو اختبار الفرضية: $H_0{:}\mu_1=\mu_2=\ldots=\mu_5$

الحل: بتطبيق الصيغ (2-13) , (2-14), (2-15) يتم حساب الآتي :

$$SST = (57)^2 + (57)^2 + \ldots + (50)^2 - (1768)^2 / 25$$
$$= 128250 - 125032.96$$
$$= 3217.04$$

$$SSBetween\ Universities = [(319)^2 + \ldots + (372)^2] /5 - (1768)^2 /25$$
$$=126424.4 - 125032.96$$
$$= 1391.44$$

$$SSe = SSWithin\ Universities = 1825.6$$

وعليه فأن جدول تحليل التباين يمكن أن يكون كما في الجدول (2-4) الآتي :

جدول (2-4) يبين تحليل التباين لبيانات التجربة (2-1)

S.O.V	d.f	S.S	M.S	F
Between groups (الجامعات)	4	1391.44	347.86	* 3.812
Within groups(Error)	20	1825.6	91.28	
Total	24	3217.04		

من جداول F (الجدول 3 في الملحق) نجد أن قيمة (F) لدرجتي حرية (20و 4) ومستوى معنوية 0.05 هي 2.87 . وبما أن قيمة (F) المستخرجة في الجدول أعلاه أكبر من قيمة (F) الجدولية فأذن نرفض الفرضية H_0 وهذا يعني أن الفروق الفردية لبين الجامعات هي فروق معنوية لذلك وضعنا العلامة (*) على قيمة (F) المستخرجة في جدول تحليل التباين للدلالة على هذه النتيجة .

سؤال// يطلب عمل جدول تحليل التباين للبيانات الناتجة عن الحالات أدناه ومقارنته مع الجدول (2-4) أعلاه.

1/ إذا تم إضافة الكمية 11 إلى كل قيمة من قيم الاستجابات للتجربة 1 .

2/ إذا تم تنقيص الكمية 11 من كل قيمة من قيم الاستجابات للتجربة 1 .

3/ إذا تم ضرب الكمية 7 بكل قيمة من قيم الاستجابات للتجربة 1 .

4/ إذا تم ضرب الكمية 1/4 بكل قيمة من قيم الاستجابات للتجربة 1 .

ملاحظة : ليس من الضروري أن يكون عدد المشاهدات في كل مجموعة متساويا ، فلو كانت بيانات التجربة (2-1) كما في الجدول (2-5) أدناه:

جدول(2-5) يبين درجات الطلبة بمادة الإحصاء

الجامعات									$y_{i.}$	n_i
الموصل	57	57	65	71	69	69	72	74	534	8
صلاح الدين	62	77	62	68					269	4
بغداد	79	95	92	69	83	90			508	6
واسط	67	57	63	70	65				322	5
ألبصره	69	79	89	85	50	60	65		497	7

$y_{..}=2130$

الحل:

$SST = (57)^2 + \ldots + (65)^2 - (2130)^2/30 = 3742$

$SSBet.(Universities) = (534)^2/8 + (269)^2/4 + (508)^2/6 + (322)^2/5 + (497)^2/7 - (2130)^2/30$

$= 1539.217$

$SSe\ (Within\ Universities) = 2202.783$

وعليه فان جدول تحليل التباين سيكون كما في الجدول (2-6) أدناه:

جدول (2-6)

تحليل التباين لبيانات الجدول 2-5 (حالة عدم تساوي التكرارات)

S.O.V	d.f	S.S	M.S	F
Between groups (الجامعات)	4	1539.217	384.804	* 4.67
Within groups(Error)	25	2202.783	88.111	
Total	29	3742		

وبما ان قيمة F المحسوبة اكبر من القيمة الجدولية $F_{0.05,4,25}=2.76$ فالفروق معنوية لبين الجامعات.