تحليل التباين : Analysis of Variance

ان فكرة تحليل التباين التي اوجدها العالم الاحصائي فيشر تعني اختبار عدة عوامل او مجموعات او عينات او مستويات .... الخ دفعة واحدة باستخدام اختبار (F) والذي هو عبارة عن نسبة تباين مجموعتين او عينتين ($\frac{S_1^2}{S_2^2}$) وبدرجات حرية (n_1-1, n_2-1)

ان تحليل التباين يعتبر من الاساليب الاحصائية المهمة من خلال ما يكشفه عن مدى تجانس او اختلاف متوسطات العينات ، ويستخدم اسلوب تحليل التباين في اختبار حول تساوي عدة اوساط حسابية بدلا من اسلوب (t) لسببين .

1- ان اختبار t يكون غير عملي كلما زاد عدد اوساط العينات وذلك لزيادة عدد الاختبارات مما يجعل العمل متعبا .
2- ان اجراء اختبار t للمقارنات المتعددة يؤدي الى احتمال الحصول على استنتاجات خاطئة بسبب صعوبة تحديد مستوى المعنوية وذلك لتداخل تلك الاوساط التي افترضناها مستقلة.

****

تحليل التباين لمعيار واحد (تجزئة التباين الى جزئين)

تتلخص هذه الطريقة بتجزئة التباين الكلي الى مجموعتين :

المجموعة الاولى التباين يعزى الى اختلاف قيم المشاهدات داخل العينة الواحدة (with in sample)

المجموعة الثانية التباين يعزى الى الاختلاف بين العينات (Between sample)

وهذه الطريقة تفترض ان العينات قيد الاختبار مستقلة فيما بينها وانها سحبت عشوائيا من مجتمعات تتوزع توزيعا طبيعيا لها التباين ذاته اي ان :

$$\sigma_1^2 = \sigma_2^2 = \ldots\ldots\ldots\ldots\ldots = \sigma_k^2 = \sigma^2$$

اي ان الانحراف المعياري متساوي ايضا

فرضية الاختبار : وتكون كالاتي :

$H_0 : \mu_1 = \mu_2 = \ldots\ldots\ldots\ldots = \mu_t$

$H_1 : \mu_1 \neq \mu_2 \neq \ldots\ldots\ldots\ldots \neq \mu_t$

t: تمثل عدد العينات قيد الاختبار

نفترض ان لدينا (t) من المجموعات او العينات ذات الاحجام $n_1, n_2, \ldots\ldots, n_t$ من المشاهدات وغالبا ما تعرف هذه العينات ب(المعاملات) في موضوع تحليل التباين ويرمز الى مفردات العينات (المشاهدات) بالرمز (Y_{ij}) ، حيث ان:

i: تشير الى رقم العينة.

j : تشير الى المفردة(المشاهدة) داخل العينة.

والجدول التالي يوضح كيفية تنظيم مشاهدات التجربة

العينات	المشاهدات				المجموع	المتوسط
	1	2	j	n	$Y_{i.}$	$\overline{Y}_{i.}$
1	Y_{11}	Y_{12}	Y_{1j}	Y_{1n}	$Y_{1.}$	$\overline{Y}_{1.}$
2	Y_{21}	Y_{22}	Y_{2j}	Y_{2n}	$Y_{2.}$	$\overline{Y}_{2.}$
3	Y_{31}	Y_{32}	Y_{3j}	Y_{3n}	$Y_{3.}$	$\overline{Y}_{3.}$
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
i	Y_{i1}	Y_{i2}	Y_{ij}	Y_{in}	$Y_{i.}$	$\overline{Y}_{i.}$
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
t	Y_{t1}	Y_{t2}	Y_{tj}	Y_{tn}	$Y_{t.}$	$\overline{Y}_{t.}$
					$Y_{..}$	$\overline{Y}_{..}$

حيث ان:

Y_{ij} : تشير الى المشاهدة j التي تنتمي الى المجموع او العينة i .

$i = 1, 2, \ldots\ldots t$; $j = 1, 2, \ldots\ldots, n$

$Y_{i.}$: تمثل مجموع مشاهدات العينة i .

$Y_{i.} = \sum_{j=1}^{ni} Y_{ij}$, i : n_i عدد المفردات ضمن المجموعة

$\overline{Y}_{i.}$: الوسط الحسابي لمشاهدات العينة i .

$$\overline{Y}_{i.} = \frac{Y_{i.}}{n_i} = \frac{\sum_{j=1}^{ni} Y_{ij}}{ni}$$

$Y_{..}$: مجموع المشاهدات في جميع العينات (المجموع العام)ويعبر عنه

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

$\overline{Y}_{..}$: تشير الى المتوسط العام لجميع المشاهدات في جميع العينات .

$$\overline{Y}_{..} = \frac{Y_{..}}{N} = \frac{\sum_{i=1}^{t}\sum_{j=1}^{ni} X_{ij}}{N}$$

حيث ان :

N: عدد المفردات (المشاهدات) في جميع العينات ويعبر عنه

$N = \sum_{i=1}^{t} n_i = n_1 + n_2 + \ldots\ldots\ldots\ldots + n_t$

ويمكن تلخيص خطوات اجراء الاختبار بالاتي :

1- حساب مجموع المربعات الكلي (SST) Total sum squares

$$SST = \sum Y_{ij}^2 - \frac{(Y_{..})^2}{N} \quad , \quad N = n.t$$

2- حساب مجموع المربعات للمعاملات (بين المجموعات)

(أ) اذا كان حجم العينات غير متساوي $SSt = \sum \frac{Y_{i.}^2}{n_i} - \frac{(Y_{..})^2}{\sum n_i}$

(ب) اذا كان حجم العينات متساوي $SSt = \sum \frac{Y_{i.}^2}{n} - \frac{(Y_{..})^2}{N}$

3- حساب مجموع المربعات داخل المعاملات ويسمى مجموع المربعات للخطأ التجريبي (SSe) sum of squares for Error

$$SSe = SST - SSt$$

4- تقسيم كل مجموع مربعات على درجات الحرية التابعة لها

5- استخراج قيمة F المحسوبة حيث انها تساوي

$$F = \frac{MSB}{MSW} = \frac{MSt}{MSe} = \frac{SSt/(t-1)}{SSe/t(n-1)}$$

6- مقارنة قيمة F المحسوبة مع قيمة F الجدولية لمستوى معنوية معين ولدرجات حرية قدرها (t-1) و [t(n-1) او N-t] ، فاذا كانت المحسوبة اكبر من الجدولية فأننا نرفض فرضية العدم H_0 ونقبل البديلة (H_1) والعكس صحيح .

قبول فرضية العدم يعني : لا توجد فروق معنوية (جوهرية)

رفض فرضية العدم وقبول البديلة: هذا يعني وجود اختلافات او فروق جوهرية بين المجموعات بشكل عام .

ويمكن ترتيب ما ذكر بجدول يطلق عليه جدول تحليل التباين ANOVA لمعيار واحد كما يلي :

مصادر التباين S . O . V	درجات الحرية d.f	مجموع المربعات S. S	متوسط المربعات M.S	قيمة الاحصاءة F
بين المعاملات (المجموعات) Between Treatments	t-1	SSt	MSB =MSt $=\frac{sst}{t-1}$	F = $\frac{MSt}{MSe}$
داخل المعاملات [Within Treatments]	t(n-1) tn-t N-t	SSe	MSw=MSe=$\frac{sse}{N-t}$	
المجموع الكلي Total	tn-1 N – 1	SST		

مثال : -

تم اخضاع مجموعة من المرضى للمعالجة بالعلاج الطبيعي وفقا لأنظمة معالجة مختلفة وعند نهاية مدة زمنية محددة تم اجراء اختبار على كل مريض لمعرفة تجانس فاعلية العلاج وكانت النتائج كالاتي :

الانظمة (العينات)	1	2	3	4	
	64	78	75	55	
	72	91	93	66	
	68	97	78	49	
	77	82	71	64	
	56	85	63	70	
	95	77	76	68	
المجموع	432	510	456	372	1770

المطلوب/هل تعتقد ان هذه البيانات تعطي دليلا كافيا لوجود فروق بين انظمة المعالجة لمستوى معين 0.05؟

الحل : -

1- فرضية الاختبار

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_1 :$ على الاقل متوسطين غير متساويين

2- احصاءة الاختبار تجري الحسابات التالية

مجموع المربعات الكلي SST = $\sum Y_{ij}^2 - \frac{(Y_{..})^2}{N}$

$$= (64)^2 + (72)^2 + (68)^2 + \cdots \ldots \ldots \ldots + (68)^2 - \frac{(1770)^2}{24} = 134192 - 130537.5$$

$$=3654.5$$

مجموع المربعات بين المجموعات(احجام العينات متساوي)

$$SSt = \sum \frac{Y_{i.}^2}{n} - \frac{(Y_{..})^2}{N}$$

$$= \frac{(432)^2+(510)^2+(456)^2+(372)^2}{6} - \frac{(1770)^2}{24} = \frac{793044}{6} - 130537$$

$$= 132174 - 130537.5 = 1635.5$$

مجموع مربعات الخطأ

$$SSe = SST - SSt$$

$$= 3654.5 - 1636.5 = 2018$$

ويمكن ترتيب النتائج في جدول تحليل التباين كالاتي :

مصادر التباين S . O . V	درجات الحرية d.f	مجموع المربعات S. S	متوسط المربعات M.S	قيمة الاحصاءة F
بين المعالجات	3	1636.5	545.5	5.406
داخل المعالجات (الخطأ)	20	2018	100.9	
المجموع الكلي Total	23	3654.5		

القرار: بمقارنة F المحسوبة (5.406) مع الجدولية لدرجات حرية (3,20) ولمستوى 5% والبالغة(3.10) نجد ان :

القيمة المحسوبة اكبر من الجدولية وعليه نرفض فرضية العدم H_0 القائلة بتساوي المتوسطات ونقبل البديلة مما يدل على ان هناك فروق معنوية بين الانظمة .

مثال 2:

تم دراسة تأثير مركبات دوائية t_1 , t_2 , t_3 , t_4 ,t_5 لعلاج احد الامراض ، وتم اختيار 24 شخص مصاب بهذا المرض(مواصفاتهم متشابهة)وتقسيمهم حسب الجدول التالي وتم اخضاع كل مجموعة لاحد هذه المركبات الدوائية وبعد ثلاثة اسابيع سجلت النتائج التي تمثل نسب الشفاء لكل مريض وكانت كما في الجدول التالي :

المجموعات المركبات الدوائية	المرضى						المجموع $Y_{i.}$	المتوسط $\bar{Y}_{i.}$
t_1	25	15	20	22	24	26	132	22
t_2	40	44	49	45	46	40	264	44
t_3	30	30	32	35			127	31.75
t_4	60	65	64	65	62		316	63.2
t_5	65	66	64				195	65
							$Y_{..} = 1034$	$\bar{Y}_{..} = 43.08$

N = 24 عدد المشاهدات الكلي

فرضية الاختبار هي :

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

H_1 : على الاقل متوسطين غير متساويين

نجري الحسابات التالية :

SST = $\sum Y_{ij}^2 - \frac{(Y_{..})^2}{N}$ مجموع المربعات الكلي

$= (25)^2 + (15)^2 + (20)^2 + \cdots \ldots \ldots \ldots + (64)^2 - \frac{(1034)^2}{24} = 51380 - 44548.166 = 6831.834$

SSt = $\sum \frac{Y_{i.}^2}{ni} - \frac{(Y_{..})^2}{N}$ مجموع المربعات بين المجموعات(احجام العينات غير متساوية)

$= \frac{(132)^2}{6} + \frac{(264)^2}{6} + \frac{(127)^2}{4} + \frac{(316)^2}{5} + \frac{(195)^2}{3} - \frac{(1034)^2}{24} = 51198.45 - 44548.166 = 6650.284$

SSe = SST - SSt مجموع مربعات الخطأ

= 6831.834 - 6650.284 = 181.55

ويمكن ترتيب النتائج في جدول تحليل التباين كالاتي :

مصادر التباين S . O . V	درجات الحرية d.f	مجموع المربعات S. S	متوسط المربعات M.S	قيمة الاحصاءة F
بين المعالجات	4	6650.284	1662.571	174.09
داخل المعالجات (الخطأ)	19	181.55	9.55	
المجموع الكلي Total	23	6831.834		

القرار: بمقارنة F المحسوبة (174.09) مع الجدولية لدرجات حرية (4,19) ولمستوى 5% والبالغة(2.9) نجد ان :

القيمة المحسوبة اكبر من الجدولية وعليه نرفض فرضية العدم H_0 ، اذن هناك فروق جوهرية بين المركبات الدوائية .