تحليل التباين
Analysis of Variance

١- الغرض من تحليل التباين

دراسة وتحليل أثر متغير أو أكثر من المتغيرات الوصفية Qualitative على متغير كمي Quantitative. ويكون من أهداف التحليل، المقارنة بين متوسطات مجموعات كل متغير من المتغيرات الوصفية محل الدراسة ، ومن الأمثلة على ذلك، يعتبر نوع المبيد متغير وصفي، فإذا كان لدينا أربعة أنواع من المبيدات، يمكن المقارنة بين متوسطات عدد الآفات المبادة باستخدام الأنواع الأربعة، لمعرفة ما إذا كانت تعطي نفس النتائج في القضاء على الآفات الضارة أم لا ، المقارنة بين نوعين من سلالات القمح ، وغير ذلك من الأمثلة، مقارنة بين عدة مناطق تنتج نوع معين من محصول معين.

فإذا كان لدينا متغير وصفي، و كانت k هي عدد مجموعات هذا المتغير، ومتوسطاتها هي: $(\mu_1, \mu_2, ..., \mu_k)$ إذا استخدمنا اختبار t في المقارنة بين كل وسطين، فإننا نحتاج إلى عدد كبير من المقارنات، وهذا يؤدي إلى تقليل احتمال قبول الفرض العدم وهو صحيح $(1-\alpha)$، ومن ثم زيادة احتمال وقوع خطأ من النوع الأول، ومن ثم يكون استخدام تحليل التباين هو التحليل المناسب في إجراء الاختبار ، حيث يمكن اختبار الفرض العدم : $H_o : \mu_1 = \mu_2 = ... = \mu_k$ مرة واحدة ، مع تقليل احتمال وقوع خطأ من النوع الأول.

أولا: تحليل التباين الأحادي
One Way ANOVA

- ***الغرض من التحليل***

هو اختبار تساوي متوسطات مجموعات Groups متغير واحد وصفي، فإذا كانت الظاهرة تحت الدراسة تشمل متغير واحد وصفي، مكون من k من المجموعات المتنافية، يمكن استخدام أبسط أنواع تحليل التباين ، وهو "تحليل التباين الأحادي"One Way ANOVA .

- ***البيانات* Data**

في هذه الحالة، تمثل المجموعات المكونة للمتغير الوصفي k من المجتمعات، ويقوم الباحث بسحب عينة عشوائية من مفردات كل مجتمع حجمها n، ثم بعد ذلك أخذ القياسات y_{ij} ، وتنظم في الشكل التالي:

Individuals	Groups (Populations)						
	1	2	----------	i	----------	k
1	y_{11}	y_{21}	----------	y_{i1}	----------	y_{k1}
2	y_{12}	y_{22}	----------	y_{i2}	----------	y_{k2}
⋮	⋮	⋮		⋮		⋮
j	y_{1j}	y_{2j}	----------	y_{ij}	----------	y_{kj}
⋮	⋮	⋮		⋮		⋮
n	y_{1n}	y_{2n}	----------	y_{in}	----------	y_{kn}

- ***الرموز المستخدمة للتعبير عن المجاميع والمتوسطات***

$: y_{ij}$	هي قيمة المشاهدة رقم j في العينة المسحوبة من المجموعة (المجتمع) رقم i ، $\{i = 1,2,..., k \quad , \quad j = 1,2,..., n\}$
$: Y_{i.}$	يمثل مجموع مشاهدات العينة رقم i ، أي أن: $Y_{i.} = \sum_{j=1}^{n} y_{ij}$
$: \bar{Y}_{i.}$	يعبر عن متوسط مشاهدات العينة رقم i ، أي أن: $\bar{Y}_{i.} = \frac{1}{n}\sum_{j=1}^{n} y_{ij} = \frac{Y_{i.}}{n}$
$: \bar{Y}_{..}$	هو الوسط الحسابي العام Grand mean، ويحسب بتطبيق المعادلة: $\bar{Y}_{..} = \frac{1}{kn}\sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij} = \frac{Y_{..}}{nk}$

• الافتراضات الخاصة بتحليل التباين الأحادي

١- المشاهدات y_{ij} مستقلة، وتتبع التوزيع الطبيعي بالخصائص التالية:

$$y_{ij} \sim \ N \ (\mu_i, \sigma_i^2)$$

٢- تباينات المجتمعات متساو، أي أن:

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

• إجراء الاختبار

قبل عرض طريقة إجراء الاختبار، فإن المقصود بتحليل التباين، هو (فصل) مكونات مجموع مربعات الانحرافات الكلية عن المتوسط العام وهو:

$$SSTo = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \overline{Y}..)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 - \frac{Y..^2}{kn}$$

$$CF = \frac{Y..^2}{kn} \qquad (1)$$

إلى مجاميع مختلفة طبقا للمصادر المسببة للاختلاف، كما تقسم درجات الحرية الكلية أيضا طبقا لهذه المصادر، وفي تحليل التباين الأحادي، يمكن تجزئة مجموع المربعات الكلية $SSTo = \sum_{i=1}^{t}\sum_{j=1}^{r}(y_{ij} - \overline{Y}_{..})^2$ إلى اثنين من المجاميع هما:

١- مجموع المربعات الذي يرجع إلى الاختلاف بين المجموعات، ويحسب بالمعادلة التالية:

$$SSB = n\sum_{i=1}^{k}(\overline{Y}_i.-\overline{Y}..)^2 = \frac{\sum_{i=1}^{k} Y_i.^2}{n} - CF \qquad (2)$$

٢- مجموع المربعات الذي يرجع إلى الاختلاف داخل المجموعات، ويحسب بالمعادلة التالية:

$$SSW = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}^2 - \overline{Y}_i.)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 - \frac{\sum_{i=1}^{k} Y_i.^2}{n} \qquad (3)$$

ومن ثم يمكن التعبير عن مجموع المربعات الكلية في الصورة التالية:

مجموع المربعات داخل المعالجات	+ مجموع المربعات بين المعالجات	= مجموع المربعات الكلية
SSW	***SSB***	***SSTo***
$\sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 - \frac{\sum_{i=1}^{k} Y_i.^2}{n}$	$\frac{\sum_{i=1}^{k} Y_i.^2}{n} - CF$	$\sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 - \frac{Y..^2}{kn}$
$df_{SSW} = k(n-1)$	$df_{SSB} = k-1$	$df_{SSTo} = kn-1$
متوسط المربعات داخل المجموعات MSW	**متوسط المربعات بين المجموعات** MSB	
$MSW = SSW/(k(n-1))$	$MSB = SSB/(k-1)$	

- ***خطوات الاختبار***

١ - ***صياغة الفروض***

$$H_o : \mu_1 = \mu_2 = ... = \mu_k$$
$$H_a : at\ \ least\ \ two\ \ unequal\ \ the\ \ other$$

٢ - ***إحصائية الاختبار***

$$F^* = \frac{MSB}{MSE} \sim \ F_{\left((k-1),\ k(n-1)\right)}$$

٣ - ***مناطق الرفض والقبول***

هي المساحة أسفل المنحنى علي يمين القيمة $F^{1-\alpha}_{\left((k-1),\ k(n-1)\right)}$

٤ - ***القرار***

إذا كان $F^* \succ F^{1-\alpha}_{\left((k-1),\ k(n-1)\right)}$ **نرفض فرض العدم ويوصى بقبول الفرض البديل.**

تطبيق (١)

رغب أحد الباحثين في اختبار تساوي متوسطات الزيادة في الوزن لثلاث أنواع من العليقة (a_1, a_2, a_3) التي يمكن أن تتناولها العجول النامية، قام باختيار ثلاث عينات عشوائية من العجول حجم كل منها 6 ، ووزع الأنواع الثلاثة من العلائق على العينات الثلاث بطريقة عشوائية ، وبعد فترة زمنية من تطبيق هذا البرنامج الغذائي قام بقياس الزيادة في الوزن بالكيلوجرام وكانت كالتالي:

a_1	a_2	a_3
14	20	9
15	22	14
16	21	21
14	27	8
16	23	7
15	19	19

بافتراض تجانس التباينات، اجري الآتي:

١- كون جدول تحليل التباين، واختبر فرض تساوي متوسطات الزيادة في الوزن للعلائق الثلاث؟ $.\alpha = 0.05$

٢- أنشأ فترة ثقة 95% للفرق بين كل وسطين (واجب منزلي).

الحل

١- تكوين جدول تحليل التباين:

بما أن : $n_1 = n_2 = n_3 = n = 6$ ، $k = 3$

حساب مجموع المربعات: SSTo، SSB، SSW **كما يلي:**

$$SSTo = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}^2 - \frac{Y..^2}{kn} = 5490 - \frac{(300)^2}{18} = 5490 - 5000 = 490$$

$$SSB = \frac{\sum_{i=1}^{k} Y_i.^2}{n} - CF = \frac{1}{6}(8100 + 17424 + 6084) - 5000 = 5268 - 5000 = 268$$

$$SSW = SSTo - SSB = 490 - 268 = 222$$

	a_1	a_2	a_3	
1	*14*	*20*	*9*	
2	*15*	*22*	*14*	
3	*16*	*21*	*21*	
4	*14*	*27*	*8*	
5	*16*	*23*	*7*	
6	*15*	*19*	*19*	
n	*6*	*6*	*6*	*18*
$Y_i.$	90	132	78	300
$Y_i.^2$	8100	17424	6084	31608
$Y.. = 300$	$\sum_{i=1}^{3}\sum_{j=1}^{6} y_{ij}^2 = 5490$			5490

تكوين جدول تحليل التباين:

S.o.V	d.f	SS	Ms	*F*
Between Groups	**2**	**268**	**134**	**9.05**
Within Groups	**15**	**222**	**14.8**	
Total	**17**	**490**		

خطوات الاختبار

صياغة الفروض:

$$H_o : \mu_1 = \mu_2 = \mu_3 = \mu$$
$$H_a : at \;\; least \;\; two \;\; unequal \;\; the \;\; other$$

إحصائية الاختبار

$$F^{*} = \frac{MSB}{MSE} = \frac{134}{14.8} = 9.05$$

المناطق الرفض والقبول

هي المساحة أسفل المنحنى علي يمين القيمة

$$F^{1-\alpha}_{\left((k-1),\, k(n-1)\right)} = F^{0.95}_{\left((2,\, 15)\right)} = 3.68$$

القرار

بما أن إذا كان $\left(F^{*} = 9.05\right) \succ \left(F^{0.95}_{\left((2,\, 15)\right)} = 3.68\right)$ **، إذا نرفض الفرض العدم**

ويستدل من ذلك أن متوسطات الزيادة في الوزن للعلائق الثلاث غير متساوية.