تصميم التجارب

Design of Experiments

الجزء الرابع

تحليل التباين

Analysis of Variance

مقدمة:

استعرضنا سابقًا اختبار الفروض التي تقارن بين متوسطي مجتمعين اثنين.

وضربنا لذلك مثالاً، حيث قمنا بمقارنة متوسطي كمية المحصول لصنفين من السماد (A و B) وذلك عن طريق اختبار الفروض الآتية:

$$H_o: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

ولكن في كثير من الأحيان نكون مهتمين بمقارنة متوسطات عدة مجتمعات (ثلاثة أو أكثر).

فعلى سبيل المثال، قد نكون مهتمين بمقارنة متوسطات كمية المحصول لأربعة أصناف من السماد (A و B و C و D).

ونرغب في معرفة هل هذه الأصناف الأربعة من السماد تعطي نفس كمية المحصول في المتوسط أم أن هناك اختلافًا معنويًا بينها.

لذا، فإننا نرغب في اختبار فرض العدم والفرض البديل الآتيين:

H_o: متوسطات كمية المحصول للأصناف الأربعة متساوية

H_1: متوسطات كمية المحصول للأصناف الأربعة غير متساوية
(يوجد على الأقل متوسطين غير متساويين)
(يوجد متوسط واحد على الأقل مختلف عن البقية)

ويمكن كتابة هذه الفروض كما يأتي:

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

H_1: يوجد متوسط واحد على الأقل (μ_i) مختلف

ملاحظة:

فرض العدم ($H_o: \mu_1 = \mu_2 = \mu_3 = \mu_4$) يعني عدم وجود فروق معنوية بين المتوسطات، أي أن المتوسطات متساوية، مما يعني عدم وجود فروق بين الأصناف الأربعة في كمية المحصول.
والفرض البديل (H_1) يعني وجود فروق معنوية بين المتوسطات، أي يوجد اختلافات بين المتوسطات (يوجد اختلاف بين اثنين على الأقل منهم)، مما يعني وجود فروق بين الأصناف الأربعة في كمية المحصول.

إن بالإمكان اختبار تساوي المتوسطات الأربعة من خلال تطبيق اختبار تي (t-test)، الذي مرمعنا سابقًا، على كل زوجين من المتوسطات.

وهذا يعني إجراء اختبار تي ست مرات، لأننا سنقوم باختبار فروض العدم والفروض البديلة الستة الآتية لكل زوج من المتوسطات على حده:

$$\begin{cases} H_o: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases} \qquad \begin{cases} H_o: \mu_1 = \mu_3 \\ H_1: \mu_1 \neq \mu_3 \end{cases} \qquad \begin{cases} H_o: \mu_1 = \mu_4 \\ H_1: \mu_1 \neq \mu_4 \end{cases}$$

$$\begin{cases} H_o: \mu_2 = \mu_3 \\ H_1: \mu_2 \neq \mu_3 \end{cases} \qquad \begin{cases} H_o: \mu_2 = \mu_4 \\ H_1: \mu_2 \neq \mu_4 \end{cases} \qquad \begin{cases} H_o: \mu_3 = \mu_4 \\ H_1: \mu_3 \neq \mu_4 \end{cases}$$

ولكن هناك عيبان لهذه الطريقة، هما:

- إن هذه الطريقة غير عملية، حيث يزداد عدد مقارنات أزواج المتوسطات بشكل كبير كلما إزداد عدد المجتمعات.
فلو كان لدينا (k) من المتوسطات، فإننا سنحتاج إلى تطبيق اختبار تي عدد من المرات مقداره:

$$\binom{k}{2} = \frac{k!}{2!(k-2)!} = \frac{k(k-1)}{2}$$

حيث يعرف مضروب العدد بالصيغة:
$$m! = m(m-1)(m-2) \ldots (2)(1)$$
وعلى سبيل المثال، لو كان لدينا ٦ من المتوسطات، فلكي نقارن جميع أزواج المتوسطات، سنحتاج إلى تطبيق اختبار تي عدد من المرات قدره:

$$\binom{6}{2} = \frac{6!}{2! \times 4!} = \frac{6 \times 5}{2} = 15$$

- وأما العيب الثاني والأبرز، فهو أن هذه الطريقة تؤدي إلى تضخم ما يسمى بمعدل الخطأ المشترك للتجربة (Experimentwise Error Rate = EWER).

ومعدل الخطأ المشترك للتجربة هو احتمال الوقوع في الخطأ من النوع الأول (Type I Error) مرة واحدة على الأقل عند إجراء اختبارات تي لجميع أزواج المتوسطات.
ومعدل الخطأ المشترك للتجربة يزداد كلما ازداد عدد المقارنات.
فعلى سبيل المثال، لوكان لدينا (k=4) من المتوسطات، وإذا ثبتنا احتمال الوقوع في الخطأ من النوع الأول (مستوى المعنوية) عند ($\alpha = 0.05$) لكل اختبار، فإن معدل الخطأ المشترك للتجربة سيكون أكبر بكثير من ($\alpha = 0.05$) حيث أنه سيكون مساويًا للمقدار:

$$EWER = 1 - (1-\alpha)^{\binom{k}{2}} = 1 - 0.95^{\binom{4}{2}} = 1 - 0.95^{6}$$

$$= 0.265$$

وهذا هو احتمال رفض فرضية عدم صحيحة واحدة على الأقل عند إجراء اختبارات تي لمقارنات جميع الأزواج.

أسلوب تحليل التباين (**An** alysis **o** f **Va** riance) (ANOVA):

إن الأسلوب الأنسب لاختبار تساوي متوسطات عدة مجتمعات هو أسلوب تحليل التباين.

ولذا، فإن أسلوب تحليل التباين يفي بالهدف الأول للتجربة والمتمثل في التحقق من وجود أو عدم وجود فروق جوهرية بين المتوسطات.

ومن خلال هذا الأسلوب نقوم باختبار الفروض الآتية باستخدام اختبار واحد فقط:

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

H_1: يوجد متوسط واحد على الأقل (μ_i) مختلف

ويتم استخدام ما يسمى باختبار إف (F - test) لإجراء هذ الاختبار.

ويُعتبر أسلوب تحليل التباين من أكثر الطرق الإحصائية استخدامًا في كثيرمن البحوث التجريبية.

وتعتمد طريقة تحليل التباين على عملية تجزيئ مجموع مربعات الانحرافات الكلي لقيم متغير الاستجابة (Y) عن متوسطه $(\sum_{i=1}^{n}(Y_i - \overline{Y})^2)$ إلى عدد من المركبات.

كل مركبة من هذه المركبات تمثل أحد المصادرالمسببة للاختلاف (Source of Variation).

وبالمثل، يتم تجزيئ درجات الحرية الكلية $(n-1)$ إلى عدد من المركبات طبقًا لمصادرالاختلاف.

مجموع مربعات الانحرافات الكلي (SS_{Total} = Total Sum of Squares):

لتعريف مجموع مربعات الانحرافات الكلي لقيم متغير الاستجابة عن متوسطها، لنفرض أن قيم متغير الاستجابة (Y) هي:

$$Y_1, Y_2, \dots, Y_n$$

ومتوسط قيم متغير الاستجابة هو:

$$\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

إن مجموع مربعات الانحرافات الكلي هو:

$$SS_{Total} = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$$
$$= \sum_{i=1}^{n} Y_i^2 - \frac{(\sum_{i=1}^{n} Y_i)^2}{n} = \sum_{i=1}^{n} Y_i^2 - \frac{Y_{\bullet}^2}{n}$$

حيث أن:

$$\sum_{i=1}^{n} Y_i^2 = Y_1^2 + Y_2^2 + \cdots + Y_n^2$$

$$Y_{\bullet} = \sum_{i=1}^{n} Y_i$$

ودرجات الحرية الكلية المصاحبة لمجموع مربعات الانحرافات الكلي (SS_{Total}) هي:

$$df_{Total} = n - 1$$

ملاحظات:

- يسمى مجموع المربعات الكلي أيضًا بمجموع المربعات الكلي المعدل (Corrected Total Sum of Squares)
- يرمز لأسلوب تحليل التباين بالرمز (ANOVA).

وتتألف خطوات إجراء تحليل التباين لاختبار تساوي المتوسطات مما يأتي:

- حساب مجموع المربعات الكلي (Total Sum of Squares).
- تحديد مصادر الاختلاف (Source of Variation).
- تجزيئ مجموع المربعات الكلي إلى عددٍ من مجاميع المربعات وفقًا لمصادر الاختلاف.
- تجزيئ درجات الحرية الكلية (Degrees of Freedom) إلى عدة أجزاء وفقًا لمصادر الاختلاف.
- حساب متوسطات المربعات (Mean of Squares).
- حساب قيمة إحصاءة اختبار إف (F-Test).
- إنشاء جدول تحليل التباين (ANOVA Table).
- اتخاذ القرار.
- عند وجود اختلاف بين المتوسطات، يتم عمل اختبارات إضافية لمقارنة المتوسطات (أو دوال في المتوسطات) لمعرفة أين تكمن الاختلافات.

تحليل التباين في إتجاه واحد (تحليل التباين للتصنيف الأحادي)

(One-Way Classification Analysis of Variance)

إن أبسط حالة لتحليل التباين هي الحالة التي يتم فيها تصنيف البيانات حسب ظاهرة واحدة (بيانات أحادية التقسيم)، حيث يكون لدينا عامل واحد له عدة مستويات، ونرغب في التحقق من وجود اختلاف بين تأثيرات هذه المستويات على متغير الاستجابة.

ففي كثيرمن التجارب يرغب الباحث في معرفة هل هناك اختلاف بين تأثير المعالجات على الوحدات التجريبية أم لا.

فيقوم الباحث بتطبيق كل معالجة على عدد من الوحدات التجريبية بشكل عشوائي (عددها $= n$)، ويسمى العدد (n) بتكرار المعالجة.

وفي نهاية التجربة يقوم الباحث بجمع المشاهدات وتصنيفها إلى مجموعات؛ كل مجموعة تمثل مشاهدات إحدى المعالجات.

ولهذا، تكون البيانات قد صُنفت على أساس خاصية واحدة (أو عامل واحد) هي نوع المعالجة.

وتُمثل كل مجموعة من البيانات عينة عشوائية مسحوبة من مجتمع إحدى المعالجات.

بعد ذلك يقوم الباحث بتحليل البيانات للتحقق من كون متوسطات المجتمعات متساوية أم لا، وذلك من خلال أسلوب تحليل التباين.

وفيما يأتي توضيحًا للعمليات الحسابية في أسلوب تحليل التباين الأحادي:

- لنفرض أن عدد من المعالجات يساوي (k) معالجة.
- ولنرمز لهذه المعالجات بالرموز: $(T_1, T_2, \dots, T_k)$.
- ونرغب في التحقق من وجود فروق بين تأثيرات هذه المعالجات على الوحدات التجريبية.
- ولنفرض أن كل معالجة من المعالجات تكررت في التجربة (n) مرة.
  أي أن كل معالجة يتم تطبيقها على (n) وحدة تجريبية بشكل عشوائي.
- لاحظ أن العدد الكلي للوحدات التجريبية يساوي:

$$N = kn$$

- يتم ترتيب بيانات الدراسة في جدول كما يأتي:

جدول بيانات التصنيف الأحادي							
المشاهدات	المعالجات (المجموعات)						
	T_1	T_2	$\cdots$	T_i	$\cdots$	T_k	
1 2 $\vdots$ j $\vdots$ n	Y_{11} Y_{12} $\vdots$ Y_{1j} $\vdots$ Y_{1n}	Y_{21} Y_{22} $\vdots$ Y_{2j} $\vdots$ Y_{2n}	$\cdots$ $\cdots$ $\vdots$ $\cdots$ $\vdots$ $\cdots$	Y_{i1} Y_{i2} $\vdots$ Y_{ij} $\vdots$ Y_{in}	$\cdots$ $\cdots$ $\vdots$ $\cdots$ $\vdots$ $\cdots$	Y_{k1} Y_{k2} $\vdots$ Y_{kj} $\vdots$ Y_{kn}	
المجموع (Total)	$Y_{1\bullet}$	$Y_{2\bullet}$	$\cdots$	$Y_{i\bullet}$	$\cdots$	$Y_{k\bullet}$	$Y_{\bullet\bullet}$ المجموع الكلي (Grand Total)
المتوسط (Mean)	$\bar{Y}_{1\bullet}$	$\bar{Y}_{2\bullet}$	$\cdots$	$\bar{Y}_{i\bullet}$	$\cdots$	$\bar{Y}_{k\bullet}$	$\bar{Y}_{\bullet\bullet}$ المتوسط الكلي (Grand Mean)

– وفيما يأتي توضيحًا للرموز في الجدول أعلاه:

الرمز	التعريف	طريقة الحساب
Y_{ij}	المشاهدة رقم (j) للمعالجة رقم (i)	Y_{ij}= تؤخذ من بيانات التجربة
i	رقم المعالجة	$i = 1, 2, \dots, k$
j	رقم المشاهدة (داخل المعالجة)	$j = 1, 2, \dots, n$
n	عدد مشاهدات المعالجة	n = عدد تكرارات المعالجة
N	عدد المشاهدات الكلي	$N = kn$
$Y_{i\bullet}$	مجموع مشاهدات المعالجة رقم (i)	$Y_{i\bullet} = \sum_{j=1}^{n} Y_{ij}$
$\bar{Y}_{i\bullet}$	متوسط مشاهدات المعالجة رقم (i)	$\bar{Y}_{i\bullet} = \frac{Y_{i\bullet}}{n} = \frac{\sum_{j=1}^{n} Y_{ij}}{n}$
$Y_{\bullet\bullet}$	المجموع الكلي (مجموع جميع المشاهدات)	$Y_{\bullet\bullet} = \sum_{i=1}^{k} Y_{i\bullet} = \sum_{i=1}^{k}\sum_{j=1}^{n} Y_{ij}$
$\bar{Y}_{\bullet\bullet}$	المتوسط الكلي (متوسط جميع المشاهدات)	$\bar{Y}_{\bullet\bullet} = \frac{Y_{\bullet\bullet}}{N} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n} Y_{ij}}{kn}$

افتراضات أسلوب تحليل التباين للتصنيف الأحادي:

لكي نتمكن من تطبيق اختبار إف (F-Test) في تحليل التباين، فلابد من توفر بعض الافتراضات الأساسية على غرار الافتراضات التي وضعنها عندما استخدمنا اختبار تي (t-Test) الخاص بمقارنة متوسطي مجتمعين.

والافتراضات هي:

- العينات مستقلة عن بعضها البعض.
- المشاهدات داخل كل عينة مستقلة عن بعضها البعض.
- العينات مسحوبة من مجتمعات طبيعية.

فمن المفترض أن يكون مجتمع قيم المعالجة رقم (i) طبيعيًا بمتوسط مقداره (μ_i) وبتباين مقداره (σ_i^2).

- تباينات المجتمعات متساوية (أي أن: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2$).

ويمكن دمج الافتراضات السابقة بالافتراض الآتي:

- المشاهدات (Y_{ij}) مستقلة عن بعضها البعض، وتتوزع وفق التوزيع الطبيعي بمتوسط (μ_i) وبتباين (σ^2)
أي أن $Y_{ij} \sim N(\mu_i , \sigma^2)$ (حيث $i = 1, 2, \dots , k$ و $j = 1, 2, \dots , n$).

الفروض المطلوب اختبارها:

نرغب في اختبار الفروض الآتية باستخدام أسلوب تحليل التباين عند مستوى الدلالة (α):

$H_o: \mu_1 = \mu_2 = \cdots = \mu_k$

$H_1:$ يوجد متوسط واحد على الأقل (μ_i) مختلف

أسلوب تحليل التباين للتصنيف الأحادي (تجزيئ مجموع المربعات الكلي):

One-Way Analysis of Variance

(Decomposition of Total Sum of Squares)

نعرف أولاً مجموع المربعات الكلي (Total Sum of Squares) على أنه مجموع مربعات انحرافات القيم (Y_{ij}) عن متوسطها العام $(\bar{Y}_{\bullet\bullet})$، أي أن:

$$SS_{Total} = \sum_{i=1}^{k} \sum_{j=1}^{n} \left(Y_{ij} - \bar{Y}_{\bullet\bullet}\right)^2$$

ودرجات الحرية المصاحبة لمجموع المربعات الكلي تساوي:

$$df_{Total} = N - 1 = kn - 1$$

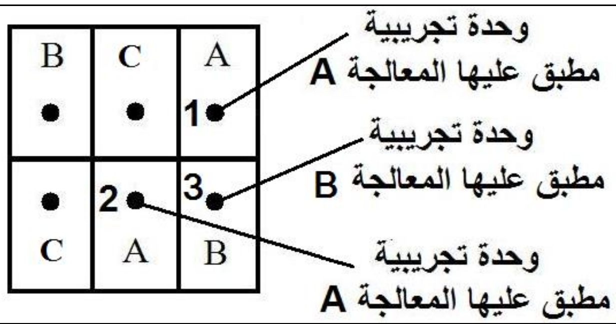
وفي تحليل التباين يتم تجزيئ مجموع المربعات الكلي (SS_{Total}) إلى مركبات وفقًا لمصادر الاختلاف.

ومصادر الاختلاف هي الأسباب التي تؤدي إلى الاختلاف في قيم متغير الاستجابة Y_{ij}.

كما يتم تجزيئ درجات الحرية المصاحبة لمجموع المربعات الكلي (df_{Total}) بالمثل إلى عدة مركبات وفقًا لمصادر الاختلاف.

ولتوضيح مفهوم مصادر الاختلاف، لنفرض أن لدينا ثلاث معالجات (A و B و C).

ولنفرض أنه تم تطبيق كل معالجة على وحدتين تجريبيتين (أي أن التكرار يساوي $n = 2$)، كما في الشكل الآتي:



نلاحظ ما يأتي من الشكل أعلاه:

- الاختلاف بين المشاهدتين ذوات الآرقام (١) و (٢) ناشئ فقط بسبب اختلاف الوحدتين (اختلافات حيوية خاص بكل وحدة) والظروف المحيطة بتطبيق المعالجة عليهما. ومصدر هذا الاختلاف يسمى بالخطأ التجريبي (أو الاختلاف داخل المعالجات).
  كما نلاحظ أنه لا يوجد هناك اختلاف بسبب اختلاف المعالجة، لأنه تم تطبيق نفس المعالجة على هاتين الوحدتين التجريبيتين.

- الاختلاف بين المشاهدتين ذوات الآرقام (١) و (٣) ناشئ من مصدرين أو بسببين؛ هما:
  (أ) بسبب اختلاف الوحدتين (اختلافات حيوية خاص بكل وحدة) والظروف المحيطة بتطبيق المعالجة عليهما (الاختلاف داخل المعالجات).
  (ب) بسبب اختلاف المعالجة، حيث تم تطبيق معالجتين مختلفتين عليهما، ومصدر هذا الاختلاف يسمى بتأثير المعالجة (أو الاختلاف بين المعالجات).

لذلك فإن لدينا مصدرين للاختلاف هما:

- اختلاف بين المعالجات (Between Treatments)
- اختلاف داخل المعالجات (الخطأ التجريبي) (Within Treatments / Experimental Error)

لذلك يتم تجزيئ مجموع المربعات الكلي (SS_{Total}) إلى مركبتين وفقًا لمصدري الاختلاف؛ وهما:

١. المركبة الأولى = مجموع المربعات بين المعالجات (بسبب المعالجات) ويرمز له بالرمز (SS_{Trt}) أو (SS_B).

(Treatment Sum of Squares / Sum of Squares due to Treatments)

٢. المركبة الثانية = مجموع المربعات داخل المعالجات (بسبب الخطأ التجريبي) ويرمز له بالرمز (SS_E) أو (SS_W).

(Error Sum of squares / Sum of Squares due to Error)

ويمكن رياضيًا إثبات الصيغة الجبرية الآتية لتفكيك مجموع المربعات الكلي إلى مركبتين (مجموع مربعات المعالجات، ومجموع مربعات الخطأ):

$$
\begin{aligned}
SS_{Total} &= \sum_{i=1}^{k}\sum_{j=1}^{n}\left(Y_{ij}-\overline{Y}_{\bullet\bullet}\right)^2 \\
&= \underbrace{n\sum_{i=1}^{k}(\overline{Y}_{i\bullet}-\overline{Y}_{\bullet\bullet})^2}_{SS_{Trt}} + \underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n}\left(Y_{ij}-\overline{Y}_{i\bullet}\right)^2}_{SS_E} \\
&= \quad SS_{Trt} \quad + \quad SS_E
\end{aligned}
$$

والجدول الآتي يوضح مجاميع المربعات ودرجات الحرية المصاحبة لها:

مجموع المربعات	صيغة التعريف الجبرية	درجات الحرية
مجموع مربعات المعالجات	$SS_{Trt} = n\sum_{i=1}^{k}(\overline{Y}_{i\bullet}-\overline{Y}_{\bullet\bullet})^2$	$df_{Trt} = k-1$
مجموع مربعات الخطأ	$SS_E = \sum_{i=1}^{k}\sum_{j=1}^{n}\left(Y_{ij}-\overline{Y}_{i\bullet}\right)^2$	$df_E = N-k = kn-k = k(n-1)$
مجموع المربعات الكلي	$SS_{Total} = \sum_{i=1}^{k}\sum_{j=1}^{n}\left(Y_{ij}-\overline{Y}_{\bullet\bullet}\right)^2$	$df_{Total} = N-1 = kn-1$

وبالمثل، يمكن رياضيًا إثبات الصيغة الجبرية الآتية لتفكيك درجات حرية مجموع المربعات الكلي إلى مركبتين (درجات حرية مجموع مربعات المعالجات، ودرجات حرية مجموع مربعات الخطأ):

$$
\begin{array}{ccccc}
\underbrace{df_{Total}} & = & \underbrace{df_{Trt}} & + & \underbrace{df_E} \\
kn-1 & = & k-1 & + & k(n-1) \\
N-1 & = & k-1 & + & (N-k)
\end{array}
$$

ولذلك، نلاحظ في الجدول السابق أن مجموع القيم في الصفين الأول والثاني تساوي القيم في الصف الثالث.

ملاحظة:

- بالرجوع إلى صيغة مجموع مربعات المعالجات:

$$SS_{Trt} = n\sum_{i=1}^{k}(\overline{Y}_{i\bullet}-\overline{Y}_{\bullet\bullet})^2$$

نلاحظ أن قيمة (SS_{Trt}) تكون كبيرة عندما تكون قيم متوسطات العينات ($\bar{Y}_{1\bullet}, \bar{Y}_{2\bullet}, \dots, \bar{Y}_{k\bullet}$) مختلفة أكثر عن بعضها البعض، وهذا مؤشرعلى وجود فروق بين المعالجات. وتكون القيمة صغيرة عندما تكون قيم متوسطات العينات قريبة من بعضها البعض، وهذا مؤشرعلى عدم وجود فروق بين المعالجات.

- بالرجوع إلى صيغة تجزيئ مجموع المربعات الكلي:

$$SS_{Total} = SS_{Trt} + SS_E$$

نلاحظ أنه عندما تكون قيمة (SS_{Trt}) كبيرة فإن قيمة (SS_E) تكون صغيرة، وهذا مؤشرعلى وجود فروق بين المعالجات.
وعندما تكون قيمة (SS_{Trt}) صغيرة فإن قيمة (SS_E) تكون كبيرة، وهذا مؤشرعلى عدم وجود فروق بين المعالجات.

- لذا، سنرى لاحقًا أن اختبار إف المستخدم لاختبار تساوي متوسطات المعالجات يعتمد أساسًا على مقارنة متوسط مجموع مربعات المعالجات (SS_{Trt}) مع متوسط مجموع مربعات الخطأ (SS_E).

صيغ حسابية لإيجاد مجاميع المربعات:

فيما يأتي صيغ حسابية لإيجاد مجاميع المربعات.

هذه الصيغ أسهل استخدامًا من صيغ التعريفات، ويمكن استخدامها مع الآلات الحاسبة بسهولة. كما أنها أكثر دقة في حال كان هناك تقريب في الحسابات.

$$N = kn$$

$CF = \frac{Y_{\bullet\bullet}^2}{N}$ (correction Factor معامل التصحيح)

مجموع المربعات	صيغة التعريف	الصيغة الحسابية
SS_{Total}	$\sum_{i=1}^{k}\sum_{j=1}^{n}\left(Y_{ij} - \bar{Y}_{\bullet\bullet}\right)^2$	$\sum_{i=1}^{k}\sum_{j=1}^{n} Y_{ij}^2 - CF$
SS_{Trt}	$n\sum_{i=1}^{k}(\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2$	$\frac{\sum_{i=1}^{k} Y_{i\bullet}^2}{n} - CF$
SS_E	$\sum_{i=1}^{k}\sum_{j=1}^{n}\left(Y_{ij} - \bar{Y}_{i\bullet}\right)^2$	$SS_{Total} - SS_{Trt}$ $\sum_{i=1}^{k}\sum_{j=1}^{n} Y_{ij}^2 - \frac{\sum_{i=1}^{k} Y_{i\bullet}^2}{n}$

متوسطات المربعات (Mean Squares):

يُعَرَّف متوسط المربعات (MS) على أنه مجموع المربعات (SS) مقسومًا على درجات الحرية المصاحبة (df)؛ أي أن:

$$MS = \frac{SS}{df}$$

فمتوسط مربعات المعالجات (Treatment Mean Squares) هو:

$$MSTrt = \frac{SS_{Trt}}{k-1} \qquad (MSB = \frac{SS_B}{k-1})$$

ومتوسط مربعات الخطأ (Error Mean Squares) هو:

$$MSE = \frac{SS_E}{k(n-1)} \qquad (MSW = \frac{SS_W}{k(n-1)})$$

إحصاءة الاختبار (اختبار إف) (F - Test):

كما ذكرنا سابقًا، فإن الهدف هو التحقق من وجود أو عدم وجود فروق معنوية بين متوسطات المعالجات عند مستوى معنوية (α) محدد.

لذا، فنحن رغب في اختبار الفروض الآتية عند مستوى الدلالة (α):

$H_o: \mu_1 = \mu_2 = \cdots = \mu_k$

$H_1:$ يوجد متوسط واحد على الأقل (μ_i) مختلف

إن إحصاءة الاختبار هي:

$$F^* = \frac{MSTrt}{MSE} \qquad (F^* = \frac{MSB}{MSW})$$

إذا كان فرض العدم صحيحًا (أي، تحت فرض العدم H_o)، فإن الإحصاءة (F^*) تتوزع وفق توزيع إف بدرجات حرية:

$$df_1 = df_{trt} = k - 1$$

$$df_2 = df_E = k(n-1)$$

أي أن:

$$F^* \sim F(k-1, k(n-1))$$

ملاحظة:

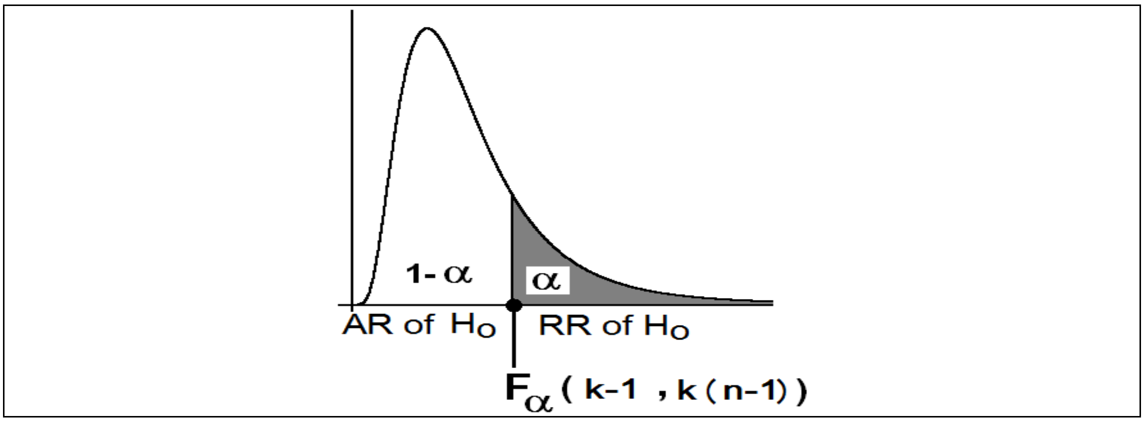
القيمة الكبيرة لإحصاءة الاختبار $(F^* = \frac{MSTrt}{MSE})$ تعني أن متوسط مربعات المعالجات $(MSTrt)$ أكبر نسبيًا من متوسط مربعات الخطأ (MSE).

وهذا مؤشر على أن الجزء الأكبر من الاختلاف ناشئ بسبب اختلاف المعالجات (وليس بسبب الخطأ أو محض الصدفة)، مما يعني عدم تساوي متوسطات المعالجات.

لذا، فإن القيمة الكبيرة لإحصاءة الاختبار تؤيد رفض فرضية العدم ($H_o: \mu_1 = \mu_2 = \cdots = \mu_k$).

منطقة رفض فرض العدم (RR) ومنطقة قبول فرض العدم (AR):

الشكل التالي يوضح منطقة رفض فرض العدم ومنطقة قبوله.



والقرار:

نظرًا لأن القيم الكبيرة لإحصاءة الاختبار تؤيد رفض فرض العدم، فإننا نرفض فرض العدم ($H_o: \mu_1 = \mu_2 = \cdots = \mu_k$) عند مستوى الدلالة ($\alpha$) إذا كانت قيمة إحصاءة الاختبار تقع في منطقة الرفض ($F^* \in RR$)، أي إذا كان:

$$F^* > F_\alpha(k-1, k(n-1))$$

ملاحظة:

إذا تم رفض فرض العدم ($H_o: \mu_1 = \mu_2 = \cdots = \mu_k$)، فإننا نستنتج أن هناك فروقًا معنوية بين المتوسطات.

وفي هذه الحالة سينصب اهتمامنا على محاولة معرفة أين تكمن تلك الفروق.

ونستطيع تحقيق ذلك من خلال ما يعرف بالمقارنات المتعددة (Multiple Comparisons) التي سوف نتناولها لاحقًا بمشيئة الله.

جدول تحليل التباين (ANOVA Table):

يمكن تلخيص حسابات أسلوب تحليل التباين في الجدول الآتي:

<table>
<tr><th>مصادر الاختلاف
Source of Variation</th><th>مجموع المربعات
Sum of Squares</th><th>درجات الحرية
Degrees of Freedom</th><th>متوسط المربعات
Mean Squares</th><th>قيمة إف
F - ratio</th></tr>
<tr><td>بين المعالجات
(Between Treatments)</td><td>SS_{Trt}</td><td>$k-1$</td><td>$MSTrt = \frac{SS_{Trt}}{k-1}$</td><td>$F^* = \frac{MSTrt}{MSE}$</td></tr>
<tr><td>داخل المعالجات (الخطأ)
(Within Treatments/ Error)</td><td>SS_E</td><td>$k(n-1)$</td><td>$MSE = \frac{SS_E}{k(n-1)}$</td><td rowspan="2">القيمة الاحتمالية

P-Value</td></tr>
<tr><td>المجموع
(Total)</td><td>SS_{Total}</td><td>$kn-1$</td><td></td></tr>
</table>

القيمة الاحتمالية (المعنوية المحسوبة) (P-Value):

تُستخدم القيمة الاحتمالية للمساعدة في اتخاذ قرار رفض أو عدم رفض فرض العدم كبديل عن استخدام القيم الجدولية ومنطقة الرفض ومنطقة القبول.

وهذه القيمة تستخدم بشكل شائع في البرامج الإحصائية.

ويمكن تعريفها بشكل مبسط على أنها أصغر قيمة لمستوى الدلالة (α) الذي يتم عندها رفض فرض العدم (H_o).

وبعبارة أدق، فهي احتمال أن تكون قيمة إحصاءة الاختبار (F^*) أكبر من القيمة المحسوبة من العينة عندما يكون فرض العدم صحيحًا (أي، تحت فرض العدم)؛ أي أن:

$$P - Value = P(F > F^* | H_o)$$

حيث أن (F) هو المتغير العشوائي الذي يتوزع وفق توزيع إف $(F(k-1, k(n-1))$.

إن القيمة الكبيرة للقيمة الاحتمالية $(P - Value)$ تؤيد فرض العدم (H_o)، بينما القيمة الصغيرة تؤيد الفرض البديل (H_1).

ويمكن صياغة القاعدة باستخدام القيمة الاحتمالية كما يأتي:

لا نرفض فرض العدم (H_o) إذا كان: $P - Value > \alpha$

نرفض فرض العدم (H_o) إذا كان: $P - Value < \alpha$

ملاحظة:

$$P-Value > \alpha \Leftrightarrow F^* < F_\alpha(k-1,k(n-1)) \quad \Leftrightarrow \quad F \in AR$$

$$P-Value < \alpha \Leftrightarrow F^* > F_\alpha(k-1,k(n-1)) \quad \Leftrightarrow \quad F \in RR$$

مثال (١):

في دراسة أُجريت لمقارنة كمية انتاج خمسة أصناف مختلفة من البرتقال (A و B وC وD و E)، قام الباحث باختيار عينة عشوائية مكونة من سبع شجرات بشكل عشوائي من كل صنف، ثم قام بتسجيل أوزان كميات الانتاج (بالكيلوجرام) وبعض الحسابات في الجدول أدناه. وكان الباحث يرغب في التحقق من وجود (أو عدم وجود) فروق معنوية بين كميات انتاج هذه الأصناف الخمسة. وقد حدد الباحث مستوى الدلالة عند ($\alpha = 0.05$).

المشاهدات	المعالجات (أصناف البرتقال)					
	$T_1 = A$	$T_2 = B$	$T_3 = C$	$T_4 = D$	$T_5 = E$	
1	7	13	20	9	18	
2	9	15	22	14	16	
3	20	16	21	21	17	
4	19	14	27	22	14	
5	11	22	23	8	22	
6	6	22	15	10	10	
7	12	16	19	7	15	
المجموع : $Y_{i\bullet}$	84	118	147	91	112	$Y_{\bullet\bullet} = 552$
المتوسط : $\bar{Y}_{i\bullet}$	12.0	16.86	21.0	13.0	16.0	$\bar{Y}_{\bullet\bullet} = 15.77$
التباين : S_i^2	30.67	13.48	13.67	38.67	13.67	$\Sigma\Sigma Y_{ij}^2 = 9720$

في هذا المثال، نلاحظ ما يأتي:

- المعالجات = أصناف البرتقال
- عدد المعالجات = k = ٥
- عدد التكرارات = n = ٧
- عدد المشاهدات الكلي = $N = kn$ = ٣٥

ولنفرض أن (μ_i) هو متوسط كمية انتاج الصنف رقم (i).

ونرغب في اختبار الفروض الآتية باستخدام أسلوب تحليل التباين عند مستوى الدلالة ($\alpha =$ 0.05):

$H_o: \mu_1 = \mu_2 = \cdots = \mu_5$

H_1: يوجد متوسط واحد على الأقل يختلف عن البقية

الحسابات الأولية:

$$\sum_{i=1}^{k}\sum_{j=1}^{n} Y_{ij}^{2} = 7^2 + 9^2 + \cdots + 10^2 + 15^2 = 9720$$

$$CF = \frac{Y_{\bullet\bullet}^{2}}{N} = \frac{552^2}{35} = 8705.83$$

$$\sum_{i=1}^{k} Y_{i\bullet}^{2} = 84^2 + 118^2 + 147^2 + 91^2 + 112^2 = 63414$$

حسابات مجامعي المربعات ودرجات الحريات ومتوسطات المربعات:

$$SS_{Total} = \sum_{i=1}^{k}\sum_{j=1}^{n} Y_{ij}^{2} - CF = 9720 - 8705.83 = 1014.17$$

$$df_{Total} = N - 1 = 35 - 1 = 34$$

$$SS_{Trt} = \frac{\sum_{i=1}^{k} Y_{i\bullet}^{2}}{n} - CF = \frac{63414}{7} - 8705.83 = 9059.14 - 8705.83$$
$$= 353.31$$

$$df_{Trt} = k - 1 = 5 - 1 = 4$$

$$MSTrt = \frac{SS_{Trt}}{k-1} = \frac{353.31}{4} = 88.328$$

$$SS_E = SS_{Total} - SS_{Trt} = 1014.17 - 353.31 = 660.86$$

$$df_E = N - k = 35 - 5 = 30$$

$$MSE = \frac{SS_E}{N-k} = \frac{660.86}{30} = 22.029$$

حساب قيمة إحصاءة الاختبار والقيمة الجدولية:

$$F^* = \frac{MSTrt}{MSE} = \frac{88.328}{22.029} = 4.01$$

$$F_\alpha\big(k-1, k(n-1)\big) = F_{0.05}(4, 30) = 2.69$$

جدول تحليل التباين:

S.O.V	SS	df	MS	F - ratio
Treatments	353.31	4	88.328	4.01
Error	660.86	30	22.029	
Total	1014.17	34		

القرار:

نظرًا لأن قيمة إحصاءة الاختبار تقع في منطقة الرفض $(F^* \in RR)$، حيث أن:

$$F^* = 4.01 \; > \; F_\alpha\big(k-1, k(n-1)\big) = 2.69$$

فإننا نرفض فرض العدم $(H_o: \mu_1 = \mu_2 = \cdots = \mu_5)$ عند مستوى الدلالة $(\alpha$=0.05$)$، ونستنتج أن هناك فروق معنوية بين متوسطات كميات انتاج أصناف البرتقال المختلفة.

ملاحظة: (حساب القيمة الاحتمالية):

يمكن التحقق من أن القيمة الاحتمالية تساوي تقريبًا:

$$P - Value \approx 0.01$$

من خلال ملاحظة أن:

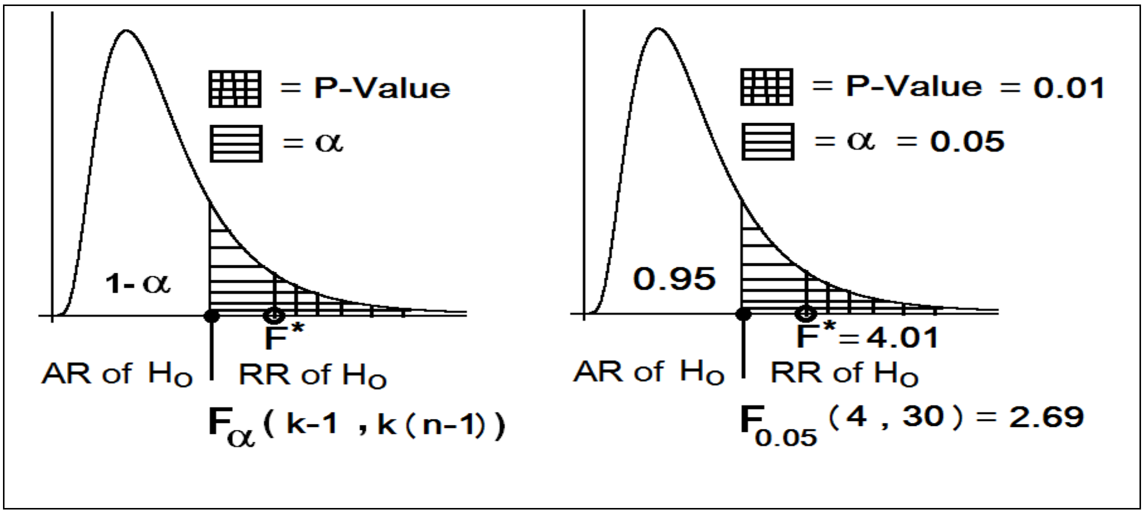
$$P - Value = P(F > F^* |\, H_o) = P(F > 4.01 \mid H_o)$$

ونظرًا لأن:

$$P - Value = 0.01 < \; \alpha = 0.05$$

فإننا نرفض فرض العدم (H_o).

والشكل التالي يوضح قيمة $(P - Value)$ مقارنة بمستوى المعنوية (α).



النموذج الخطي لتحليل التباين الأحادي:

:(Linear Model for One-Way ANOVA)

النموذج الخطي هو معادلة رياضية تفسر مصادراختلاف مشاهدات قيم متغير الاستجابه (Y_{ij}) عن بعضها البعض. ويعطى بالصيغة الآتية:

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

$$\mathrm{j} = 1, 2, \dots, \mathrm{n} \;\; ; \;\; \mathrm{i} = 1, 2, \dots, \mathrm{k}$$

حيث أن (Y_{ij}) هي المشاهدة رقم (j) للمعالجة (أو للمجموعة) رقم (i)، وتتكون من ثلاث مركبات؛ وهي:

١. μ = المتوسط العام؛ وهو قيمة ثابتة ومجهولة.
٢. α_i = تأثير المعالجة رقم (i) وهي قيمة مجهولة.
وقد تكون ثابتة أو عشوائية، ويعتمد ذلك على نوع التصميم، كما سنرى لاحقًا.
٣. ϵ_{ij} = مركبة الخطأ العشوائي للمشاهدة رقم (j) للمعالجة رقم (i).

والافتراضات الأساسية للنموذج الخطي لتحليل التباين الأحادي هي:

١. إذا كانت التأثيرات (α_i) ثابتة، فلا بد أن تحقق الشرط: $\sum_{i=1}^{k} \alpha_i = 0$.
وأما إن كانت عشوائية، فلا بد أن تكون متوزعة وفق التوزيع الطبيعي بمتوسط (0) وتباين (σ_α^2)، أي أن $\alpha_i \sim \mathrm{N}(0\,, \sigma_\alpha^2)$، وتكون مستقلة عن بعضها البعض، وتكون أيضًا مستقلة عن مركبة الخطأ العشوائي (ϵ_{ij}).
٢. الأخطاء العشوائية (ϵ_{ij}) تتوزع وفق التوزيع الطبيعي بمتوسط (0) وتباين (σ_i^2)، أي أن $\epsilon_{ij} \sim \mathrm{N}(0\,, \sigma_i^2)$، وتكون مستقلة عن بعضها البعض.
٣. التباينات تكون متساوية، أي أن: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2$.

ومن الافتراضات السابقة، يمكن التحقق مما يأتي حول توزيع قيم متغير الاستجابة (Y_{ij}):

١. المتوسط (أو التوقع الرياضي Expected Value) للمشاهدة (Y_{ij}) هو:

$$E(Y_{ij}) = \mu_i$$
$$= \mu + \alpha_i$$

٢. تباين المشاهدة (Y_{ij}) هو:

$$Var(Y_{ij}) = \sigma_i^2$$

٣. توزيع المشاهدة (Y_{ij}) هو التوزيع الطبيعي.
٤. المشاهدات (Y_{ij}) مستقلة عن بعضها البعض.

ويتضح مما سبق أن المشاهدات (Y_{ij}) تتوزع وفق التوزيع الطبيعي بمتوسط $(\mu + \alpha_i)$ وتباين (σ_i^2)، وتكون مستقلة عن بعضها البعض ، أي أن: $Y_{ij} \sim N(\mu + \alpha_i \,, \sigma_i^2)$.

وعند تبني فرض تساوي التباينات $(\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2)$، يصبح:

$$Y_{ij} \sim N(\mu + \alpha_i \,, \sigma^2)$$

ملاحظة:

هدفنا الأساس هو التحقق من صحة فرض العدم القائل بأن متوسطات المعالجات متساوية، أي:

$$H_o: \mu_1 = \mu_2 = \cdots = \mu_k$$

ونظرًا لأن:

$$\mu_i = \mu + \alpha_i$$

فإن الفرض السابق يمكن صياغتة بشكل مكافئ على النحو الآتي:

$$H_o: \alpha_1 = \alpha_2 = \cdots = \alpha_k$$

وعند وضع الشرط $\sum_{i=1}^{k} \alpha_i = 0$ (لنموذج التأثيرات الثابتة)، يصبح فرض العدم على الشكل الآتي:

$$H_o: \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$$

لاحظ أن الفرض $(H_o: \mu_1 = \mu_2 = \cdots = \mu_k)$ يعني عدم وجود فروق بين المتوسطات، مما يعني عدم وجود اختلاف بين تأثيرات المعالجات على متغير الاستجابة.

وأما الفرض $(H_o: \alpha_1 = \alpha_2 = \cdots = \alpha_k)$ فيعني أن تأثيرت المعالجات متساوية، مما يعني عدم وجود اختلاف في تأثيرات المعالجات على متغير الاستجابة.

لذا، فإن الفرضين لهما نفس المعنى.

الأساس النظري لإحصاءة إختبار تحليل التباين (اختبار إف):

يمكن كتابة متوسطات المربعات (MSE) و $(MSTrt)$ كما يأتي:

(١) متوسط مربعات المعالجات:

$$MSTrt = \frac{SS_{Trt}}{k-1} = \frac{n \sum_{i=1}^{k} (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2}{k-1} = n \left(\frac{\sum_{i=1}^{k} (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2}{k-1} \right)$$

ولتحليل هذا المقدار، بناءً على النظرية الإحصائية، فمن المعلوم أن متوسط العينة $(\overline{\mathrm{Y}}_{\mathrm{i}\bullet})$ المسحوبة من التوزيع الطبيعي $N(\mu_i\,,\sigma^2)$، يتوزع وفق التوزيع الطبيعي بمتوسط يساوي $(E(\overline{\mathrm{Y}}_{\mathrm{i}\bullet})=\mu_i)$ وتباين يساوي $(Var(\overline{\mathrm{Y}}_{\mathrm{i}\bullet})=\sigma^2/n)$، أي أن:

$$\overline{\mathrm{Y}}_{\mathrm{i}\bullet}\sim \mathrm{N}(\mu_i\,,\frac{\sigma^2}{\mathrm{n}})$$

وإذا كانت العينات مستقلة عن بعض البعض، وكانت متوسطات التوزيعات متساوية $(\mu_1=\mu_2=\cdots=\mu_k=\mu)$، فإن متوسطات العينات:

$$\overline{\mathrm{Y}}_{1\bullet}\,,\overline{\mathrm{Y}}_{2\bullet}\,,\ldots,\overline{\mathrm{Y}}_{\mathrm{k}\bullet}$$

تشكل عينة عشوائية من التوزيع الطبيعي $\mathrm{N}(\mu\,,\sigma^2/n)$. لذلك فإن تباين عينة المتوسطات الذي يعطى بالصيغة الآتية:

$$\frac{\sum_{\mathrm{i}=1}^{\mathrm{k}}(\overline{\mathrm{Y}}_{\mathrm{i}\bullet}-\overline{\mathrm{Y}}_{\bullet\bullet})^2}{k-1}$$

يمثل تقديرًا منصفًا (أو غير متحيز/غير منحاز) (Unbiased) لتباين توزيع المتوسطات (σ^2/n).

ولذلك فإن المقدار $(\sum_{\mathrm{i}=1}^{\mathrm{k}}(\overline{\mathrm{Y}}_{\mathrm{i}\bullet}-\overline{\mathrm{Y}}_{\bullet\bullet})^2\,/(\mathrm{k}-1))$ يكون منصفًا للمقدار (σ^2/n).

وبالتالي، يكون المقدار الآتي:

$$n\left(\frac{\sum_{\mathrm{i}=1}^{\mathrm{k}}(\overline{\mathrm{Y}}_{\mathrm{i}\bullet}-\overline{\mathrm{Y}}_{\bullet\bullet})^2}{k-1}\right)=\frac{n\sum_{\mathrm{i}=1}^{\mathrm{k}}(\overline{\mathrm{Y}}_{\mathrm{i}\bullet}-\overline{\mathrm{Y}}_{\bullet\bullet})^2}{k-1}=\frac{SS_{Trt}}{k-1}=\ MSTrt$$

يمثل تقديرًا منصفًا (Unbiased) للتباين المشترك (σ^2).

أي أن متوسط مربعات المعالجات $(MSTrt)$ يكون مقدرًا منصفًا للتباين المشترك (σ^2)، فقط عندما تكون متوسطات المعالجات متساوية، أي عندما يكون فرض العدم $(H_o:\mu_1=\mu_2=\cdots=\mu_k)$ صحيحًا.

(٢) متوسط مربعات الخطأ:

$$MSE=\frac{SS_E}{k(n-1)}=\frac{\sum_{\mathrm{i}=1}^{\mathrm{k}}\sum_{\mathrm{j}=1}^{\mathrm{n}}\left(\mathrm{Y}_{\mathrm{ij}}-\overline{\mathrm{Y}}_{i\bullet}\right)^2}{k(n-1)}=\frac{\sum_{\mathrm{i}=1}^{\mathrm{k}}\left(\sum_{\mathrm{j}=1}^{\mathrm{n}}\left(\mathrm{Y}_{\mathrm{ij}}-\overline{\mathrm{Y}}_{i\bullet}\right)^2\right)}{k(n-1)}$$

$$=\frac{\sum_{\mathrm{i}=1}^{\mathrm{k}}\left(\frac{\sum_{\mathrm{j}=1}^{\mathrm{n}}\left(\mathrm{Y}_{\mathrm{ij}}-\overline{\mathrm{Y}}_{i\bullet}\right)^2}{(n-1)}\right)}{k}=\frac{\sum_{\mathrm{i}=1}^{\mathrm{k}}\mathrm{S}_{\mathrm{i}}^2}{k}$$

حيث أن (S_i^2) هو تباين العينة رقم (i)، أي أن:

$$S_i^2 = \frac{\sum_{j=1}^{n}\left(Y_{ij} - \overline{Y}_{i\bullet}\right)^2}{n-1}$$

ولتحليل المقدار $(MSE = \sum_{i=1}^{k} S_i^2 /k)$، بناءً على النظرية الإحصائية، فمن المعلوم أن تباين العينة (S_i^2) هو تقدير منصف لتباين المجتمع المشترك (σ^2) دائمًا بغض النظر عن صحة أو عدم صحة فرض العدم $(H_o: \mu_1 = \mu_2 = \cdots = \mu_k)$.

ولذلك فإن متوسط تباينات العينات $(S_1^2\,, S_2^2\,, \ldots\ \,, S_k^2)$ المعطى بالصيغة الآتية:

$$\frac{\sum_{i=1}^{k} S_i^2}{k} = MSE$$

هو أيضًا تقدير منصف لتباين المجتمع المشترك (σ^2)

مما سبق، نستنتج ما يأتي:

(١) يعتبر المقدار $MSE = SS_E/k(n-1)$ تقديرًا منصفًا لتباين المجتمع المشترك (σ^2) دائمًا.

(٢) يعتبر المقدار $MSTrt = SS_{Trt}\ /(k-1)$ تقديرًا منصفًا لتباين المجتمع المشترك (σ^2) فقط عندما يكون فرض العدم $(H_o: \mu_1 = \mu_2 = \cdots = \mu_k)$ صحيحًا.

ومن هذا الاستنتاج، نرى أن قيم المقدارين $MSTrt$ و MSE تكون قريبة من بعضها البعض عندما يكون فرض العدم صحيحًا؛ وذلك لأنهما يقدران نفس الكمية (σ^2).

ولذلك فإن إحصاءة اختبار تحليل التباين، المعطاة بالصيغة الآتية:

$$F^* = \frac{MSTrt}{MSE}$$

تعتمد على مقارنة المقدارين MSE و $MSTrt$.

إضافةً إلى ما سبق، فإن مربعات المتوسطات MSE و $MSTrt$ هي متغيرات عشوائية. ويمكن التحقق من أن متوسطاتها (توقعاتها الرياضية) هي:

$$E(MSTrt) = \sigma^2 + \frac{n\sum_{i=1}^{k}\alpha_i^2}{k-1}$$

$$E(MSE) = \sigma^2$$

ومن هنا يتضح بشكل أوضح دور إحصاءة الاختبار $(F^* = MSTrt\ /MSE)$ في الحكم على صحة أو عدم صحة فرض العدم:

$$H_o: \mu_1 = \mu_2 = \cdots = \mu_k$$
$$\Leftrightarrow$$
$$H_o: \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$$

حيث يلاحظ أنه إذا كان فرض العدم صحيحًا، فإن ($\sum_{i=1}^{k} \alpha_i^2 = 0$)، وبالتالي فإن:

$$E(MSTrt) = E(MSE) = \sigma^2$$

وهذا يؤدي إلى أن القيمة المتوقعة الإحصاءة ($F^* = MSTrt\ /MSE$) تكون قريبة من الواحد الصحيح.

وأما إذا كانت فرضية العدم غير صحيحية، فسيكون ($\sum_{i=1}^{k} \alpha_i^2 > 0$)، وهذا يؤدي إلى أن:

$$E(MSTrt) > E(MSE)$$

وبالتالي سيكون من المتوقع أن تكون قيمة إحصاءة الاختبار (F^*) كبيرة (أكبر من الواحد).

ولذلك فإننا نرفض فرض العدم عندما تكون قيمة إحصاءة الاختبار (F^*) كبيرة (أي أكبر من القيمة الجدولية ($F_\alpha\big(k-1, k(n-1)\big)$).

الخطأ المعياري (Standard Error) لمتوسط عينة:

يُعطى الخطأ المعياري لمتوسط العينة ($\overline{\mathrm{Y}}_{i\bullet}$) بالصيغة الآتية:

$$\mathrm{S}_{\overline{\mathrm{Y}}_{i\bullet}} = \sqrt{\frac{\mathrm{MSE}}{\mathrm{n}}}$$

الخطأ المعياري (Standard Error) للفرق بين متوسطي عينتين:

يُعطى الخطأ المعياري للفرق بين متوسطي عينتين ($\overline{\mathrm{Y}}_{i\bullet} - \overline{\mathrm{Y}}_{k\bullet}$) بالصيغة الآتية:

$$\mathrm{S}_{\overline{\mathrm{Y}}_{i\bullet}-\overline{\mathrm{Y}}_{k\bullet}} = \sqrt{\frac{\mathrm{MSE}}{\mathrm{n}} + \frac{\mathrm{MSE}}{\mathrm{n}}} = \sqrt{\mathrm{MSE}\left(\frac{1}{\mathrm{n}} + \frac{1}{\mathrm{n}}\right)} = \sqrt{\frac{2\mathrm{MSE}}{\mathrm{n}}}$$

مثال:

في مثال (١) وجدنا أن: $n = 7$ و $MSE = 22.029$.

لذلك فإن الخطأ المعياري لمتوسط العينة ($\overline{\mathrm{Y}}_{i\bullet}$) يساوي:

$$S_{\bar{Y}_{i\bullet}} = \sqrt{\frac{22.029}{7}} = 1.774$$

والخطأ المعياري للفرق بين متوسطي عينتين $(\bar{Y}_{i\bullet} - \bar{Y}_{k\bullet})$ يساوي:

$$S_{\bar{Y}_{i\bullet}-\bar{Y}_{k\bullet}} = \sqrt{\frac{2(22.029)}{7}} = 2.5088$$

تحليل التباين للتصنيف الأحادي مع عدم تساوي حجوم العينات:

One-Way Analysis of Variance with Unequal Sample Sizes

لنفرض أن حجم العينة المأخوذة من مجتمع المعالجة رقم (i) هو (n_i).

فيكون لدينا عينات المعالجات كما في الجدول الآتي:

جدول بيانات التصنيف الأحادي (حالة عدم تساوي حجوم العينات)

المشاهدات	المعالجات (المجموعات)						
	T_1	T_2	$\cdots$	T_i	$\cdots$	T_k	
1	Y_{11}	Y_{21}	$\cdots$	Y_{i1}	$\cdots$	Y_{k1}	
2	Y_{12}	Y_{22}	$\cdots$	Y_{i2}	$\cdots$	Y_{k2}	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
j	Y_{1j}	Y_{2j}	$\cdots$	Y_{ij}	$\cdots$	Y_{kj}	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$\vdots$	Y_{1n_1}	$\vdots$	$\cdots$	Y_{in_i}	$\cdots$	$\vdots$	
n_i		Y_{2n_2}	$\vdots$		$\vdots$	Y_{kn_k}	
المجموع (Total)	$Y_{1\bullet}$	$Y_{2\bullet}$	$\cdots$	$Y_{i\bullet}$	$\cdots$	$Y_{k\bullet}$	$Y_{\bullet\bullet}$ المجموع العام (Grand Total)
المتوسط (Mean)	$\bar{Y}_{1\bullet}$	$\bar{Y}_{2\bullet}$	$\cdots$	$\bar{Y}_{i\bullet}$	$\cdots$	$\bar{Y}_{k\bullet}$	$\bar{Y}_{\bullet\bullet}$ المتوسط العام (Grand Mean)

ونظرًا لأن لدينا (k) من المعالجات، فستكون حجوم العينات هي:

$$n_1 \,,\, n_2 \,, \ldots \,,\, n_k$$

ويكون العدد الكلي للمشاهدات هو:

$$N = \sum_{i=1}^{k} n_i$$

ولتحليل التباين في حالة عدم تساوي حجوم العينات، فإننا نستخدم نفس الأسلوب السابق والمتبع في الحالة التي تكون فيها حجوم العينات متساوية، فيما عدا عمل بعض التعديلات على صيغ بعض القوانين الحسابية التي تستخدم لحساب المجاميع والمتوسطات ومجاميع المربعات ودرجات الحريات.

والتعديلات هي على النحو الآتي:

$$\mathrm{Y}_{\bullet\bullet} = \sum_{\mathrm{i}=1}^{\mathrm{k}} \sum_{\mathrm{j}=1}^{n_i} \mathrm{Y}_{\mathrm{ij}} \qquad , \qquad \bar{\mathrm{Y}}_{\bullet\bullet} = \frac{\mathrm{Y}_{\bullet\bullet}}{\mathrm{N}}$$

$$\mathrm{Y}_{i\bullet} = \sum_{\mathrm{j}=1}^{n_i} \mathrm{Y}_{\mathrm{ij}} \qquad , \qquad \overline{\mathrm{Y}}_{i\bullet} = \frac{\mathrm{Y}_{i\bullet}}{n_i}$$

$$\mathrm{CF} = \frac{{\mathrm{Y}_{\bullet\bullet}}^2}{\mathrm{N}}$$

$$SS_{Total} = \sum_{\mathrm{i}=1}^{\mathrm{k}} \sum_{\mathrm{j}=1}^{\mathrm{n_i}} \left(\mathrm{Y}_{\mathrm{ij}} - \overline{\mathrm{Y}}_{\bullet\bullet}\right)^2 = \sum_{\mathrm{i}=1}^{\mathrm{k}} \sum_{\mathrm{j}=1}^{\mathrm{n_i}} {\mathrm{Y}_{\mathrm{ij}}}^2 - \mathrm{CF}$$

$$df_{Total} = N - 1$$

$$SS_{Trt} = \sum_{\mathrm{i}=1}^{\mathrm{k}} n_i \left(\overline{\mathrm{Y}}_{\mathrm{i}\bullet} - \overline{\mathrm{Y}}_{\bullet\bullet}\right)^2 = \sum_{\mathrm{i}=1}^{\mathrm{k}} \frac{{\mathrm{Y}_{\mathrm{i}\bullet}}^2}{n_i} - CF$$

$$df_{Trt} = k - 1$$

$$SS_E = \sum_{\mathrm{i}=1}^{\mathrm{k}} \sum_{\mathrm{j}=1}^{n_i} \left(\mathrm{Y}_{\mathrm{ij}} - \overline{\mathrm{Y}}_{i\bullet}\right)^2 = SS_{Total} - SS_{Trt}$$

$$df_E = N - k$$

ويبقى تجزيئ مجموع المربعات الكلي وتجزيئ درجات الحرية المصاحبة له صحيحًا كما سبق، أي أن:

$$SS_{Total} = SS_{Trt} + SS_E$$

$$df_{Total} = df_{Trt} + df_E$$

ويتم كذلك إنشاء جدول تحليل التباين بنفس الطريقة المتبعة في حالة حجوم العينات المتساوية.

ويتم حساب إحصاءة الاختبار (F^*) بنفس الطريقة.

ونرفض فرض العدم عند مستوى الدلالة (α) إذا كان:

$$F^* > F_\alpha(k-1, N-k)$$

مثال ٢:

أجريت تجربة لدراسة تأثير ثلاثة أنواع من الأدوية المختلفة على مستوى الكوليسترول في دم الأرانب. واستخدم في هذه التجربة ١٨ أرنبًا. وزعت الأرانب على ثلاث مجموعات بطريقة عشوائية، كل مجموعة تحوي ٦ أرانب، واستلمت كل مجموعة نوعًا واحدًا من الأدوية الثلاثة بشكل عشوائي. وخلال هذه التجربة نفق ثلاثة أرانب قبل أخذ قياساتها. وكانت قياسات مادة الكوليسترول على بقية الأرانب مع بعض الحسابات كالآتي:

	المعالجات (أنواع الدواء)			
	$T_1 = D_1$	$T_2 = D_2$	$T_3 = D_3$	
المشاهدات	15 22 17 16	16 20 19 22 25 21	17 13 16 18 12	
حجم العينة: n_i	4	6	5	
المجموع : $Y_{i\bullet}$	70	123	76	الجموع العام $Y_{\bullet\bullet} = 269$
المتوسط : $\bar{Y}_{i\bullet}$	17.5	20.5	15.2	المتوسط العام $\bar{Y}_{\bullet\bullet} = 17.933$
التباين : S_i^2	9.667	9.1	6.7	مجموع المربعات $\Sigma\Sigma Y_{ij}^2 = 5003$

وكان الهدف من الدراسة هو التحقق من وجود أو عدموجود فروق بين مستويات مادة الكوليسترول للأدوية الثلاثة، عند مستوى الدلالة عند $(\alpha = 0.05)$.

في هذا المثال، نلاحظ ما يأتي:

- المعالجات = أنواع الأدوية
- عدد المعالجات = k = 3
- عدد التكرارات: $n_1 = 4,\ n_2 = 6,\ n_3 = 5$
- عدد المشاهدات الكلي: $N = \sum_{i=1}^{3} n_i = 4 + 6 + 5 = 15$
- ولنفرض أن (μ_i) هو متوسط مستوى مادة الكوليسترول لنوع الدواء رقم (i).

ونرغب في اختبار الفروض الآتية باستخدام أسلوب تحليل التباين عند مستوى الدلالة $(\alpha = 0.05)$:

$H_o: \mu_1 = \mu_2 = \mu_3$

H_1: يوجد متوسط واحد على الأقل يختلف عن البقية

الحسابات الأولية:

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij}^2 = 15^2 + 22^2 + \cdots + 18^2 + 12^2 = 5003$$

$$CF = \frac{Y_{\bullet\bullet}^2}{N} = \frac{269^2}{15} = 4824.0667$$

حسابات مجاميع المربعات ودرجات الحريات ومتوسطات المربعات:

$$SS_{Total} = \sum_{\mathrm{i}=1}^{\mathrm{k}} \sum_{\mathrm{j}=1}^{n_i} \mathrm{Y_{ij}}^2 - \mathrm{CF} = 5003 - 4824.0667 = 178.9333$$

$$df_{Total} = N - 1 = 15 - 1 = 14$$

$$SS_{Trt} = \sum_{\mathrm{i}=1}^{\mathrm{k}} \frac{\mathrm{Y_{i\bullet}}^2}{n_i} - \mathrm{CF} = \left(\frac{70^2}{4} + \frac{123^2}{6} + \frac{76^2}{5}\right) - 4824.0667$$

$$= 4901.7 - 4824.0667 = 77.63333$$

$$df_{Trt} = k - 1 = 3 - 1 = 2$$

$$MSTrt = \frac{SS_{Trt}}{k-1} = \frac{77.63333}{2} = 38.8167$$

$$SS_E = SS_{Total} - SS_{Trt} = 178.9333 - 77.63333 = 101.3$$

$$df_E = N - k = 15 - 3 = 12$$

$$MSE = \frac{SS_E}{N-k} = \frac{101.3}{12} = 8.4417$$

حساب قيمة إحصاءة الاختباروالقيمة الجدولية:

$$F^* = \frac{MSTrt}{MSE} = \frac{38.8167}{8.4417} = 4.5982$$

$$F_\alpha(k-1, N-k) = F_{0.05}(2,12) = 3.89$$

جدول تحليل التباين:

S.O.V	SS	df	MS	F - ratio
Treatments	77.63333	2	38.8167	4.5982
Error	101.3	12	8.4417	
Total	178.9333	1٤		

القرار:

نظرًا لأن قيمة إحصاءة الاختبار تقع في منطقة الرفض ($F^* \in RR$)، حيث أن:

$$F^* = 4.5982 > F_\alpha(k-1, N-k) = 3.89$$

فإننا نرفض فرض العدم $(H_o: \mu_1 = \mu_2 = \mu_3)$ عند مستوى الدلالة $(\alpha=0.05)$، ونستنتج أن هناك فروق معنوية بين متوسطات كميات مادة الكوليسترول لأنواع الأدوية المختلفة.

ملاحظة:

يمكن التحقق من أن القيمة الاحتمالية (القيمة المعنوية) تساوي:

$$P - Value = 0.0329$$

ونظرًا لأن هذه القيمة أقل من قيمة مستوى الدلالة $(\alpha=0.05)$، فإننا نرفض فرض العدم.

مخالفات افتراضات تحليل التباين (Violations of ANOVA Assumptions):

ذكرنا سابقًا جميع الافتراضات الأساسية الضرورية لإجراء تحليل التباين.

وكان من ضمن تلك الافتراضات اشتراط تجانس تباينات مجتمعات المعالجات:

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

حيث يُعد هذا الافتراض من أهم افتراضات تحليل التباين.

لذلك، فلا بد من التحقق من هذا الشرط قبل إجراء تحليل التباين، لأن عدم تحقق هذا الشرط يؤدي إلى عدم صحة اختبار إف وبالتالي يقودنا إلى استنتاج غير دقيق.

اختبار تجانس التباينات (اختبار بارتليت) (Bartlett's Test):

نرغب في اختبار الفروض الآتية:

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

H_1: يوجد تباين واحد (σ_i^2) على الأقل يختلف عن البقية

وتجدر الإشارة إلى أنه في الحالة الخاصة عندما يكون هناك تباينين اثنين فقط، أي عند الرغبة في اختبار الفروض الآتية:

$$H_o: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

فيمكن استخدام اختبار إف (F-Test) الذي ذكرناه في فصل سابق.

ففي تلك الحالة كنا نرفض فرض العدم عند مستوى الدلالة (α) إذا كان:

$$F^* < F_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1) \quad \text{أو} \quad F^* > F_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)$$

ونقبل فرض العدم إذا كان:

$$F_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1) < F^* < F_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)$$

حيث أن:

$$F^* = \frac{S_1^2}{S_2^2}$$

والآن سنتستعرض اختبارًا إحصائيًا لاختبار تجانس التباينات للحالة العامة، أي عندما يكون لدينا أكثر من مجتمعين، ونرغب في اختبار الروض:

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

H_1: يوجد تباين واحد $\left(\sigma_i^2\right)$ على الأقل يختلف عن البقية

لنفرض أن تباينات العينات هي:

$$\mathrm{S}_1^2 , \mathrm{S}_2^2 , \dots \ , \mathrm{S}_\mathrm{k}^2$$

ولنفرض أن حجوم العينات هي:

$$n_1 , n_2 , \dots , n_k$$

نعرف إحصاءة اختبار بارتليت على النحو الآتي:

$$V^* = \frac{\lambda}{C}$$

حيث أن:

$$\lambda = (N-k)\ln\left(S_p^2\right) - \sum_{i=1}^{k}(n_i - 1)\ln(\mathrm{S}_\mathrm{i}^2)$$

$$C = 1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k}\frac{1}{n_i - 1} - \frac{1}{N-k}\right)$$

$$S_p^2 = \frac{\sum_{\mathrm{i}=1}^{\mathrm{k}}(n_i - 1)S_i^2}{N-k} = MSE$$

$ln = log_e =$ اللوغاريم الطبيعي

يلاحظ ما يأتي حول قيمة (λ):

- عندما تكون تباينات العينات (S_i^2) مختلفة عن بعضها البعض (غير متجانسة)، فإن قيمة (λ) تكون كبيرة، وبالتالي تكون قيمة إحصاءة الاختبار (V^*) كبيرة.

- عندما تكون تباينات العينات (S_i^2) متساوية، فإن قيمة (λ) تساوي الصفر، وبالتالي تكون قيمة إحصاءة الاختبار (V^*) مساوية للصفر.

إن التوزيع التقريبي لإحصاءة الاختبار (V^*) تحت فرض العدم هو توزيع مربع كاي (χ^2) بدرجات حرية مقدارها $(df = \nu = k-1)$، أي أن:

$$V^* \sim \chi^2(k-1)$$

ونظرًا لأن القيمة الكبيرة لإحصاءة الاختبار تؤيد عدم تجانس تباينات العينات، فإننا نرفض فرض العدم عند مستوى الدلالة (α) إذا كان:

$$V^* > \chi^2_\alpha(k-1)$$

مثال (٣):

نرغب في اختبار تجانس تباينات المجتمعات في مثال (٢) عند مستوى الدلالة عند ($\alpha = 0.05$).

الجدول الآتي يلخص بعض الحسابات التي وجدناها في مثال (٢):

	المعالجات (أنواع الدواء)		
	$T_1 = D_1$	$T_2 = D_2$	$T_3 = D_3$
حجم العينة: n_i	4	6	5
التباين : S_i^2	9.667	9.1	6.7

عدد المعالجات (الجتمعات) هو: $k = 3$

عدد المشاهدات هو: $N = \sum_{i=1}^{3} n_i = 4 + 6 + 5 = 15$

ونرغب في اختبار الفروض الآتية عند مستوى الدلالة ($\alpha = 0.05$):

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

$H_1:$ يوجد تباين واحد (σ_i^2) على الأقل يختلف عن البقية

ولإيجاد قيمة إحصاءة الاختبار نحسب ما يأتي:

$$S_p^2 = \frac{\sum_{\mathrm{i=1}}^{\mathrm{k}} (n_i - 1) S_i^2}{N - k} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 + (n_3 - 1)S_3^2}{N - k}$$

$$= \frac{(4-1)(9.667) + (6-1)(9.1) + (5-1)(6.7)}{15 - 3}$$

$$= \frac{101.301}{12}$$

$$= 8.44175 \qquad (= MSE = MSW)$$

كما يمكن حساب (S_p^2) من جدول تحليل التباين حيث أن:

$$S_p^2 = MSE$$

$$\lambda = (N-k)\ln\left(S_p^2\right) - \sum_{i=1}^{k}(n_i - 1)\ln(\mathrm{S}_{\mathrm{i}}^2)$$
$$= (N-k)\ln\left(S_p^2\right)$$
$$-\left[(n_1-1)\ln(S_1^2) + (n_2-1)\ln(S_2^2) + (n_3-1)\ln(S_3^2)\right]$$
$$= (15-3)\ln(8.44175)$$
$$-\left[(4-1)\ln(9.667) + (6-1)\ln(9.1) + (5-1)\ln(6.7)\right]$$
$$= 25.59828 - 25.45596$$
$$= 0.14232$$

$$C = 1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k}\frac{1}{n_i-1} - \frac{1}{N-k}\right)$$

$$= 1 + \frac{1}{3(3-1)}\left[\left(\frac{1}{4-1} + \frac{1}{6-1} + \frac{1}{5-1}\right) - \frac{1}{15-3}\right]$$

$$= 1 + \frac{1}{3(3-1)}(0.783333 - 0.083333)$$

$$= 1 + \frac{1}{3(3-1)}(0.7)$$

$$= 1.11667$$

وبالتعويض في صيغة إحصاءة الاختبار، نجد أن القيمة تساوي:

$$V^* = \frac{\lambda}{C} = \frac{0.14232}{1.11667} = 0.12745$$

وأما القيمة الجدولية (القيمة الحرجة) فهي:

$$\chi_\alpha^2(k-1) = \chi_{0.05}^2(2) = 5.99$$

ونظرًا لأن قيمة إحصاءة الاختبار أصغر من القيمة الجدولية:

$$V^* < \chi_\alpha^2(k-1)$$

فإننا لا نرفض فرض العدم القائل بتساوي التباينات عند مستوى الدلالة ($\alpha = 0.05$). وبناءً على ذلك، فإن افتراض تساوي التباينات لتحليل التباين يعتبر متحقق.

تحليل التباين الثنائي (للنموذج الجمعي):

Two-Way Analysis of Variance for (Additive Model)

في بعض الحالات يكون الباحث مهتمًا بدراسة تأثير عاملين على متغيرالاستجابة.

وفي هذه الحالة يتم تصنيف البيانات في اتجاهين.

فعلى سبيل المثال، قد يكون الباحث مهتمًا بدراسة تأثير نوع السماد (العامل الأول (A)) وكمية الماء (العامل الثاني (B)) على كمية انتاج أحد أصناف القمح (متغير الاستجابة Y).

ولدراسة هذا النوع من التصاميم، لنفرض أن عدد مستويات العامل الأول (A) يساوي (a)، وعدد مستويات العامل الثاني (B) يساوي (b).

ولذلك سيكون لدينا عدد مقداره (ab) من الخلايا التصالبة التي تشكل مستويات العاملين معًا.

وفي هذا الجزء، سنفرض أيضًا أن هناك مشاهدة واحدة فقط (عدد التكرارات = ١) لكل خلية، ولذلك سيكون عدد المشاهدات (N=ab).

كما سنفرض عدم وجود تداخل بيني (أو تفاعل بيني) (Interaction Effect) بين العاملين. وبمعنى آخر، علاقة أي عامل بمتغير الاستجابة لا تعتمد على مستويات العامل الآخر؛ أي لا يوجد تأثير لأي عامل على علاقة متغير الاستجابة مع العامل الآخر.

وإذا فرضنا أن مستويات العامل الأول (A) هي: $A_1\ ,\ A_2\ ,\dots\ ,\ A_a$

ومستويات العامل الثاني (B) هي: $B_1\ ,\ B_2\ ,\dots\ ,\ B_a$

فإن تراكيب المعالجات هي:

		مستويات العامل الثاني (B)			
		B_1	B_2	$\cdots$	B_b
مستويات العامل الأول (A)	A_1	A_1B_1	A_1B_2	$\cdots$	A_1B_b
	A_2	A_2B_1	A_2B_2	$\cdots$	A_2B_b
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	A_a	A_aB_1	A_aB_2	$\cdots$	A_aB_b

حيث (A_iB_j) تعني المعالجة التي يتم فيها تطبيق المستوى (A_i) للعامل الأول والمستوى (B_j) للعامل الثاني.

ويتم تلخيص بيانات التصنيف الثنائي كما في الجدول الآتي:

		مستويات العامل الثاني (B)				المجموع	المتوسط
		B_1	B_2	$\cdots$	B_b	$Y_{i\bullet}$	$\bar{Y}_{i\bullet}$
مستويات العامل الأول (A)	A_1	Y_{11}	Y_{12}	$\cdots$	Y_{1b}	$Y_{1\bullet}$	$\bar{Y}_{1\bullet}$
	A_2	Y_{21}	Y_{22}	$\cdots$	Y_{2b}	$Y_{2\bullet}$	$\bar{Y}_{2\bullet}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	A_a	Y_{a1}	Y_{a2}	$\cdots$	Y_{ab}	$Y_{a\bullet}$	$\bar{Y}_{a\bullet}$
المجموع $Y_{\bullet j}$		$Y_{\bullet 1}$	$Y_{\bullet 2}$	$\cdots$	$Y_{\bullet b}$	المجموع العام $Y_{\bullet\bullet}$	
المتوسط $\bar{Y}_{\bullet j}$		$\bar{Y}_{\bullet 1}$	$\bar{Y}_{\bullet 2}$	$\cdots$	$\bar{Y}_{\bullet b}$	المتوسط العام $\bar{Y}_{\bullet\bullet}$	

وفيما يأتي تعريف بالمجاميع والمتوسطات المذكورة في الجدول السابق:

المجموع / المتوسط	طريقة الحساب	التعريف
$Y_{i\bullet}$	$\sum_{j=1}^{b} Y_{ij}$	مجموع مشاهدات المستوى رقم (i) للعامل الأول
$\bar{Y}_{i\bullet}$	$\frac{Y_{i\bullet}}{b}$	متوسط مشاهدات المستوى رقم (i) للعامل الأول
$Y_{\bullet j}$	$\sum_{i=1}^{a} Y_{ij}$	مجموع مشاهدات المستوى رقم (j) للعامل الثاني
$\bar{Y}_{\bullet j}$	$\frac{Y_{\bullet j}}{a}$	متوسط مشاهدات المستوى رقم (j) للعامل الثاني
$Y_{\bullet\bullet}$	$\sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij}$	المجموع العام
$\bar{Y}_{\bullet\bullet}$	$\frac{Y_{\bullet\bullet}}{N}$	المتوسط العام

النموذج الخطي لتحليل التباين (للنموذج الجمعي):

كما ذكرنا سابقًا، فإن النموذج الخطي هو معادلة رياضية تفسر مصادراختلاف مشاهدات قيم متغير الاستجابه (Y_{ij}) عن بعضها البعض.

ونظرًا لأننا قد فرضنا عدم وجود تداخل بيني (Interaction Effect) بين العاملين. فيمكن كتابة النموذج الخطي الذي يتم من خلاله تجزيئ كل مشاهدة إلى عدة مركبات كما يأتي:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

$$j = 1, 2, \dots, b \;\; ; \;\; i = 1, 2, \dots, a$$

حيث أن (Y_{ij}) هي مشاهدة المعالجة المكونة من المستوى رقم (i) للعامل الأول (A) والمستوى رقم (j) للعامل الثاني (B)، وتتكون من أربع مركبات؛ وهي:

١. μ = المتوسط العام؛ وهو قيمة ثابتة ومجهولة.
٢. α_i = تأثير المستوى رقم (i) للعامل الأول وهي قيمة مجهولة.

٣. β_j = تأثير المستوى رقم (j) للعامل الثاني وهي قيمة مجهولة.
٤. ϵ_{ij} = مركبة الخطأ العشوائي المصاحبة لمشاهدة المعالجة المكونة من المستوى رقم (i) للعامل الأول والمستوى رقم (j) للعامل الثاني.

ويفترض أن تتوزع الأخطاء العشوائية (ϵ_{ij}) بشكل مستقل وفق التوزيع الطبيعي بمتوسط (0) وتباين (σ^2)، أي أن $\epsilon_{ij} \sim \mathrm{N}(0\,,\sigma^2)$.

وللنموذج ذي التأثيرات الثابتة، يفترض أن يكون:

$$\sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = 0$$

بناءً على النموذج السابق والافتراضات السابقة، فإن متوسط (القيمة المتوقعة) للمشاهدة (Y_{ij}) هو:

$$\begin{aligned} E(Y_{ij}) &= \mu_{ij} \\ &= \mu + \alpha_i + \beta_j \end{aligned}$$

ومن هذه الصيغة، نرى أن متوسط المشاهدة (Y_{ij}) مكون من مجموع المتوسط العام (μ) وتأثير العامل الأول (α_i) وتأثير العامل الثاني (β_j)؛ ولذلك يُقال بأن النموذج جمعيًا (Additive Model).

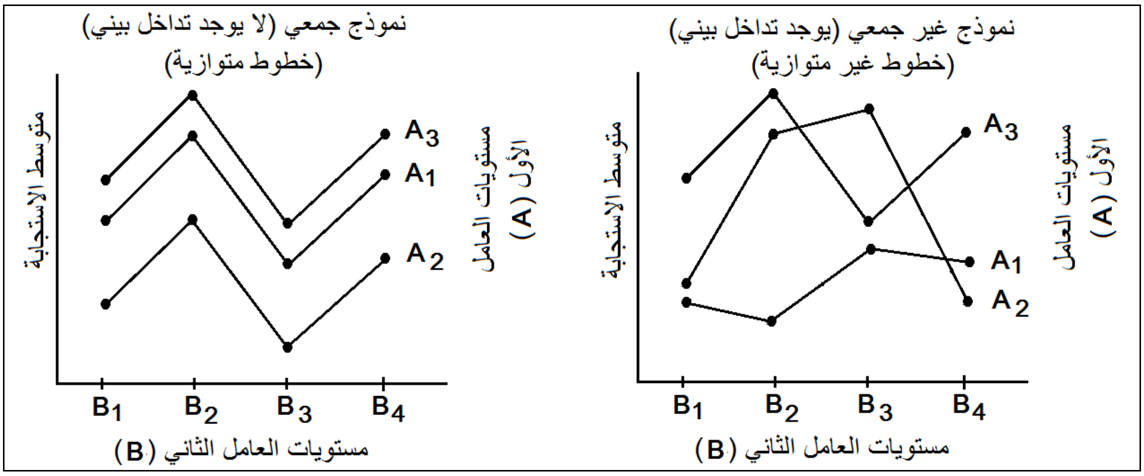
ويلاحظ في هذا النموذج الجمعي أن الفرق بين متوسطي مستويين لأحد العوامل هو نفسه لجميع مستويات العامل الآخر. وبعبارة أخرى، فإن الفرق بين متوسطي مستويين لأحد العوامل لا يتغيير بتغير مستويات العامل الآخر. أي أن طبيعة العلاقة بين متغير الاستجابة مع أحد العوامل لا تتأثر بالعامل الآخر. ولذلك، نقول بأنه لا يوجد تداخل بيني بين العاملين.

فعلى سبيل المثال، لو ثبتنا العامل الثاني عند المستوى رقم (٣) وغيرنا مستويات العامل الأول من المستوى رقم (1) إلى المستوى رقم (2) لكان متوسط الفرق بين المشاهدتين يساوي:

$$\begin{aligned} E(Y_{13}) - E(Y_{23}) &= (\mu + \alpha_1 + \beta_3) - (\mu + \alpha_2 + \beta_3) \\ &= \alpha_1 - \alpha_2 \end{aligned}$$

ومن هنا، نلاحظ أن هذا الفرق مستقل عن مستوى العامل الثاني.

والشكل الآتي يوضح حالتين؛ أحدهما لنموذج جمعي (لا يوجد تداخل بيني بين العاملين)، والنموذج الآخر غير جمعي (يوجد تداخل بيني بين العاملين).



الفروض المراد اختبارها (Hypotheses):

نرغب في معرفة ما يأتي في تحليل التباين الثنائي للنموذج الجمعي:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

(١) هل هناك فروق بين مستويات العامل الأول (A). وبعبارة أخرى، هل يؤثر العامل الأول على متغير الاستجابة.
وتتم الإجابة على هذا التساؤل من خلال اختبار الفروض الآتية:

$$\mathrm{H_o^A}: \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$$
$\mathrm{H_1^A}: \alpha_i \neq 0$ لقيمة واحدة على الأقل

(٢) هل هناك فروق بين مستويات العامل الثاني (B). وبعبارة أخرى، هل يؤثر العامل الثاني على متغير الاستجابة.
وتتم الإجابة على هذا التساؤل من خلال اختبار الفروض الآتية:

$$\mathrm{H_o^B}: \beta_1 = \beta_2 = \cdots = \beta_b = 0$$
$\mathrm{H_1^B}: \beta_j \neq 0$ لقيمة واحدة على الأقل

وسوف نستخدم أسلوب تحليل التباين للإجابة على هذه التساؤلات.

تحليل التباين:

كما مر معنا سابقًا، فإننا نعرف مجموع المربعات الكلي (Total Sum of Squares) على أنه مجموع مربعات انحرافات القيم ($\mathrm{Y_{ij}}$) عن متوسطها العام ($\bar{Y}_{\bullet\bullet}$)، أي أن:

$$SST = SS_{Total} = \sum_{i=1}^{a} \sum_{j=1}^{b} \left(\mathrm{Y_{ij}} - \bar{\mathrm{Y}}_{\bullet\bullet}\right)^2$$

ودرجات الحرية المصاحبة لمجموع المربعات الكلي تساوي:

$$df_{Total} = N - 1 = ab - 1$$

ويتم تجزيئ مجموع المربعات الكلي (SS_{Total}) إلى مركبات وفقًا لمصادر الاختلاف كما يأتي:

١. مجموع مربعات العامل الأول (A):
(مركبة تأثير العامل الأول / الاختلاف بسبب العامل الأول)

$$SSA = \sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 = b\sum_{i=1}^{a}(\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2$$

ودرجات الحرية المصاحبة لمجموع مربعات العامل الأول تساوي:

$$df_A = a - 1$$

ومتوسط مجموع مربعات العامل الأول هو:

$$MSA = \frac{SSA}{df_A} = \frac{SSA}{a-1}$$

٢. مجموع مربعات العامل الثاني (B):
(مركبة تأثير العامل الثاني / الاختلاف بسبب العامل الثاني)

$$SSB = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet}\right)^2 = a\sum_{j=1}^{b}\left(\bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet}\right)^2$$

ودرجات الحرية المصاحبة لمجموع مربعات العامل الثاني تساوي:

$$df_B = b - 1$$

ومتوسط مجموع مربعات العامل الثاني هو:

$$MSB = \frac{SSB}{df_B} = \frac{SSB}{b-1}$$

٣. مجموع مربعات الخطأ:
(مركبة الخطأ التجريبي)

$$SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(Y_{ij} - \bar{Y}_{i\bullet} - \bar{Y}_{\bullet j} + \bar{Y}_{\bullet\bullet}\right)$$

ودرجات الحرية المصاحبة لمجموع مربعات الخطأ تساوي:

$$df_E = (a-1)(b-1)$$

ومتوسط مجموع مربعات الخطأ هو:

$$MSE = \frac{SSE}{df_E} = \frac{SSE}{(a-1)(b-1)}$$

ويمكن التحقق من العلاقات الجبرية الاتية لتجزيئة مجمع المربعات الكلي ودرجات الحرية المصاحبة وفقًا لمصادر الاختلاف:

$$SS: \quad SST = SSA + SSB + SSE$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\left(Y_{ij} - \overline{Y}_{\bullet\bullet}\right)^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})^2$$

$$+ \sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet}\right)^2$$

$$+ \sum_{i=1}^{a}\sum_{j=1}^{b}\left(Y_{ij} - \overline{Y}_{i\bullet} - \overline{Y}_{\bullet j} + \overline{Y}_{\bullet\bullet}\right)$$

$$df: \quad df_T = df_A + df_B + df_E$$

$$ab - 1 = (a-1) + (b-1) + (a-1)(b-1)$$

صيغ حسابية لإيجاد مجاميع المربعات:

فيما يأتي صيغ حسابية لإيجاد مجاميع المربعات.

هذه الصيغ أسهل استخدامًا من صيغ التعريفات، ويمكن استخدامها مع الآلات الحاسبة بسهولة. كما أنها أكثر دقة في حال كان هناك تقريب في الحسابات.

$N = ab$

$CF = \frac{Y_{\bullet\bullet}^2}{ab}$ (correction Factor معامل التصحيح)

$SST = \sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij}^2 - CF$

$SSA = \frac{1}{b}\sum_{i=1}^{a} Y_{i\bullet}^2 - CF$

$SSB = \frac{1}{a}\sum_{j=1}^{b} Y_{\bullet j}^2 - CF$

$SSE = SST - SSA - SSB$

إحصاءات الاختبارات (اختبارات إف) (F - Tests):

نستعرض فيما يأتي طريقة اختبارات الفروض التي نهتم بها:

(١) من أجل اختبار أهمية (معنوية) تأثير العامل الأول (A)، أي من أجل اختبار الفروض الآتية:

$$H_o^A: \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$$
$$H_1^A: \alpha_i \neq 0 \quad \text{لقيمة واحدة على الأقل}$$

فإننا نستخدم إحصاءة الاختبار الآتية:

$$F^A = \frac{MSA}{MSE}$$

إذا كان فرض العدم صحيحًا، فإن الإحصاءة (F^A) تتوزع وفق توزيع إف بدرجات حرية ($df_1 = df_A = a - 1$) و ($df_2 = df_E = (a-1)(b-1)$)، أي أن:

$$F^A \sim F(a-1, (a-1)(b-1))$$

ونرفض فرض العدم عند مستوى الدلالة (α) إذا كان:

$$F^A > F_\alpha(a-1, (a-1)(b-1))$$

(٢) ومن أجل اختبار أهمية (معنوية) تأثير العامل الثاني (B)، أي من أجل اختبار الفروض الآتية:

$$H_o^B: \beta_1 = \beta_2 = \cdots = \beta_b = 0$$
$$H_1^B: \beta_j \neq 0 \quad \text{لقيمة واحدة على الأقل}$$

فإننا نستخدم إحصاءة الاختبار الآتية:

$$F^B = \frac{MSB}{MSE}$$

إذا كان فرض العدم صحيحًا، فإن الإحصاءة (F^B) تتوزع وفق توزيع إف بدرجات حرية ($df_1 = df_B = b - 1$) و ($df_2 = df_E = (a-1)(b-1)$)، أي أن:

$$F^B \sim F(b-1, (a-1)(b-1))$$

ونرفض فرض العدم عند مستوى الدلالة (α) إذا كان:

$$F^B > F_\alpha(b-1, (a-1)(b-1))$$

ملاحظة:

إذا تم رفض أحد فروض العدم (H_o^A أو H_o^B)، فإننا نستنتج أن هناك فروقًا معنوية بين المتوسطات. وفي هذه الحالة سينصب اهتمامنا على محاولة معرفة أين تكمن تلك الفروق. ونستطيع تحقيق ذلك من خلال ما يعرف بالمقارنات المتعددة (Multiple Comparisons) التي سوف نتناولها لاحقًا بمشيئة الله.

جدول تحليل التباين (ANOVA Table):

يمكن تلخيص حسابات أسلوب تحليل التباين الثنائي (للنموذج الجمعي) في الجدول الآتي:

مصادر الاختلاف Source of Variation	مجموع المربعات Sum of Squares	درجات الحرية Degrees of Freedom	متوسط المربعات Mean Squares	قيم إف F - ratios
العامل الأول (A) (Levels of A)	SSA	$a-1$	MSA	$F^A = \frac{MSA}{MSE}$
العامل الأول (B) (Levels of B)	SSB	$b-1$	MSB	$F^B = \frac{MSB}{MSE}$
الخطأ (Error)	SSE	$(a-1)(b-1)$	MSE	
المجموع (Total)	SST	$ab-1$		

التوقعات الرياضية لمتوسطات المربعات:

إن متوسطات المربعات (MSA و MSB و MSE) هي متغيرات عشوائية.

ويمكن التحقق من أن متوسطاتها (توقعاتها الرياضية) هي:

$$E(MSA) = \sigma^2 + \frac{b\sum_{i=1}^{a}\alpha_i^2}{a-1}$$

$$E(MSB) = \sigma^2 + \frac{a\sum_{j=1}^{b}\beta_j^2}{b-1}$$

$$E(MSE) = \sigma^2$$

ومن هنا يتضح بشكل أوضح دور إحصاءات الاختبار (F^A و F^B) في الحكم على صحة أو عدم صحة فروض العدم.

حيث يلاحظ أنه إذا كان فرض العدم الخاص بالعامل الأول ($\mathrm{H_o^A}: \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$) صحيحًا، فإن ($\sum_{i=1}^{a}\alpha_i^2 = 0$)، وبالتالي فإن:

$$E(MSA) = E(MSE) = \sigma^2$$

وهذا يؤدي إلى أن القيمة المتوقعة للإحصاءة $(F^A = MSA / MSE)$ تكون قريبة من الوحد الصحيح. وأما إذا كان فرض العدم $(\mathrm{H}_\mathrm{o}^\mathrm{A})$ غير صحيح، فسيكون $(\sum_{i=1}^{a} \alpha_i^2 > 0)$، وهذا يؤدي إلى أن:

$$E(MSA) > E(MSE)$$

وبالتالي سيكون من المتوقع أن تكون قيمة إحصاءة الاختبار (F^A) كبيرة (أكبر من الواحد).

ولذلك فإننا نرفض فرض العدم $(\mathrm{H}_\mathrm{o}^\mathrm{A})$ عندما تكون قيمة إحصاءة الاختبار (F^A) كبيرة (أي أكبر من القيمة الجدولية $\big(F_\alpha\big(a-1,(a-1)(b-1)\big)$.

وبالمثل يمكن أن يُقال حول إحصاءة الاختبار (F^B) الخاصة بالعامل الثاني.

الخطأ المعياري لمتوسطات العينات والخطأ المعياري للفروق بين متوسطات العينات:

بالنسبة لمستويات العامل الأول:

الخطأ المعياري لمتوسط عينة أحد مستويات العامل الأول $(\overline{\mathrm{Y}}_{i\bullet})$ هو:

$$\mathrm{S}_{\overline{\mathrm{Y}}_{i\bullet}} = \sqrt{\frac{\mathrm{MSE}}{b}}$$

والخطأ المعياري للفرق بين متوسطي عينتين من مستويين للعامل الأول $(\overline{\mathrm{Y}}_{\mathrm{i}\bullet} - \overline{\mathrm{Y}}_{\mathrm{k}\bullet})$ هو:

$$\mathrm{S}_{\overline{\mathrm{Y}}_{i\bullet}-\overline{\mathrm{Y}}_{k\bullet}} = \sqrt{\frac{2\mathrm{MSE}}{b}}$$

بالنسبة لمستويات العامل الثاني:

الخطأ المعياري لمتوسط عينة أحد مستويات العامل الثاني $(\overline{\mathrm{Y}}_{\bullet\mathrm{j}})$ هو:

$$\mathrm{S}_{\overline{\mathrm{Y}}_{\bullet j}} = \sqrt{\frac{\mathrm{MSE}}{a}}$$

والخطأ المعياري للفرق بين متوسطي عينتين من مستويين للعامل الثاني $(\overline{\mathrm{Y}}_{\bullet j} - \overline{\mathrm{Y}}_{\bullet\mathrm{m}})$ هو:

$$\mathrm{S}_{\overline{\mathrm{Y}}_{\bullet\mathrm{j}}-\overline{\mathrm{Y}}_{\bullet\mathrm{m}}} = \sqrt{\frac{2\mathrm{MSE}}{a}}$$

مثال (٤):

في أحد البحوث المتعلقة بدراسة تأثير حركة المرور على تلوث الهواء في إحدى المدن، أُخذت عينات من ٥ أماكن مختلفة (A_1, A_2, A_3, A_4, A_5) وفي ٤ أشهر مختلفة (B_1, B_2, B_3, B_4)، وقيست في تلك العينات كمية المواد الصلبة، وكانت البيانات كما في الجدول أدناه.

وكان الباحث مهتمًا بمعرفة ما يأتي:

١. هل هناك فرق في درجة التلوث بين الأماكن الخمسة.
٢. هل هناك فرق في درجة التلوث بين الأزمنة الأربعة.
٣. إيجاد الخطأ المعياري لمتوسط عينة أحد مستويات العامل الأول.
٤. إيجاد الخطأ المعياري للفرق بين متوسطي عينتين لمستويات العامل الأول.
٥. إيجاد الخطأ المعياري لمتوسط عينة أحد مستويات العامل الثاني.
٦. إيجاد الخطأ المعياري للفرق بين متوسطي عينتين لمستويات العامل الثاني.

		مستويات العامل الثاني - الشهر (B)				المجموع $Y_{i\bullet}$	المتوسط $\bar{Y}_{i\bullet}$
		B_1 أكتوبر	B_2 يناير	B_3 مايو	B_4 سبتمبر		
مستويات العامل الأول - المكان (A)	A_1	76	82	68	63	289	72.25
	A_2	67	69	59	56	251	62.75
	A_3	81	96	67	64	308	77.00
	A_4	56	59	54	58	227	56.75
	A_5	51	70	42	37	200	50.00
المجموع $Y_{\bullet j}$		331	376	290	278	المجموع العام $Y_{\bullet\bullet} = 1275$	
المتوسط $\bar{Y}_{\bullet j}$		66.2	75.2	58.0	55.6	المتوسط العام $\bar{Y}_{\bullet\bullet} = 63.75$	
						مجموع المربعات $\sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij}^2$ =84853	

نلاحظ ما يلي:

العامل الأول (A) هو المكان وعدد مستوياته (a=5)

العامل الثاني (B) هو الزمان وعدد مستوياته (b=4)

يوجد مشاهدة واحدة لكل خلية.

حساب مكونات جدول تحليل التباين:

$N = ab = 5\times4=20$

$$\mathrm{CF} = \frac{\mathrm{Y}_{\bullet\bullet}^{\ 2}}{\mathrm{ab}} = \frac{1275^2}{20} = 81281.25$$

$$SST = \sum_{\mathrm{i}=1}^{\mathrm{a}} \sum_{\mathrm{j}=1}^{\mathrm{b}} \mathrm{Y}_{\mathrm{ij}}^{\ 2} - \mathrm{CF} = 84853 - 81281.25 = 3571.75$$

$$df_T = ab - 1 = 20 - 1 = 19$$

$$SSA = \frac{1}{b} \sum_{\mathrm{i}=1}^{a} \mathrm{Y}_{\mathrm{i}\bullet}^{\ 2} - \mathrm{CF} = \frac{1}{4}(\ 289^2 +\ 251^2 +\ 308^2 + 227^2\ + 200^2\ ) - 81281.25 = \frac{1}{4}(332915) - 81281.25 = 1947.50$$

$$df_A = a - 1 = 5 - 1 = 4$$

$$MSA = \frac{SSA}{a-1} = \frac{1947.50}{4} = 486.875$$

$$SSB = \frac{1}{a} \sum_{\mathrm{j}=1}^{b} \mathrm{Y}_{\bullet j}^{\ 2} - \mathrm{CF} = \ \frac{1}{5}(\ 331^2 +\ 376^2 +\ 290^2 + 278^2\ ) - 81281.25 = \frac{1}{5}(412321) - 81281.25 = 1182.95$$

$$df_B = b - 1 = 4 - 1 = 3$$

$$MSB = \frac{SSB}{b-1} = \frac{1182.95}{3} = 394.3167$$

$$SSE = SST - SSA - SSB = 3571.75 - 1947.50 - 1182.95 = 441.30$$

$$df_E = (a-1)(b-1) = 4 \times 3 = 12$$

$$MSE = \frac{SSE}{(a-1)(b-1)} = \frac{441.30}{12} = 36.775$$

$$F^A = \frac{MSA}{MSE} = \frac{486.875}{36.775} = 13.239$$

$$F_{0.05}(df_A, df_E) = F_{0.05}(4, 12) = 3.26$$

$$F^B = \frac{MSB}{MSE} = \frac{394.3167}{36.775} = 10.722$$

$$F_{0.05}(df_B, df_E) = F_{0.05}(3, 12) = 3.49$$

ونقوم بتلخيص حسابات تحليل التباين في الجدول الآتي:

مصادر الاختلاف S.O.V	مجموع المربعات SS	درجات الحرية df	متوسط المربعات MS	قيم إف F - ratios
العامل الأول (A) (Levels of A)	1947.50	4	486.875	F^A=13.239
العامل الأول (B) (Levels of B)	1182.95	3	394.3167	F^B=10.722
الخطأ (Error)	441.30	12	36.775	
المجموع (Total)	3571.75	19		

(١) نظرًا لأن (F^A>3.26)، فإننا نرفض فرض العدم القائل بتساوي مستويات العامل الأول:

$$\mathrm{H}_{\mathrm{o}}^{\mathrm{A}}: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$$

ونستنتج أن هناك فروق معنوية في كمية التلوث بين الأماكن الخمسة.

(٢) نظرًا لأن (F^B>3.49)، فإننا نرفض فرض العدم القائل بتساوي مستويات العامل الثاني:

$$\mathrm{H}_{\mathrm{o}}^{\mathrm{B}}: \beta_1 = \beta_2 = \beta_2 = \beta_4 = 0$$

ونستنتج أن هناك فروق معنوية في كمية التلوث بين الأزمنة الأربعة.

(٣) الخطأ المعياري لمتوسط عينة أحد مستويات العامل الأول ($\bar{\mathrm{Y}}_{i\bullet}$) هو:

$$S_{\bar{\mathrm{Y}}_{i\bullet}} = \sqrt{\frac{\mathrm{MSE}}{b}} = \sqrt{\frac{36.775}{4}} = 3.0321$$

(٤) الخطأ المعياري للفرق بين متوسطي عينتين من مستويين للعامل الأول ($\bar{\mathrm{Y}}_{\mathrm{i}\bullet} - \bar{\mathrm{Y}}_{\mathrm{k}\bullet}$) هو:

$$\mathrm{S}_{\bar{\mathrm{Y}}_{i\bullet}-\bar{\mathrm{Y}}_{k\bullet}} = \sqrt{\frac{2\mathrm{MSE}}{b}} = \sqrt{\frac{(2)(36.775)}{4}} = 4.2881$$

(٥) الخطأ المعياري لمتوسط عينة أحد مستويات العامل الثاني ($\bar{\mathrm{Y}}_{\bullet\mathrm{j}}$) هو:

$$\mathrm{S}_{\bar{\mathrm{Y}}_{\bullet j}} = \sqrt{\frac{\mathrm{MSE}}{a}} = \sqrt{\frac{36.775}{5}} = 2.7120$$

(٦) الخطأ المعياري للفرق بين متوسطي عينتين من مستويين للعامل الثاني ($\bar{\mathrm{Y}}_{\bullet j} - \bar{\mathrm{Y}}_{\bullet\mathrm{m}}$) هو:

$$\mathrm{S}_{\bar{\mathrm{Y}}_{\bullet\mathrm{j}}-\bar{\mathrm{Y}}_{\bullet\mathrm{m}}} = \sqrt{\frac{2\mathrm{MSE}}{a}} = \sqrt{\frac{(2)(36.775)}{5}} = 3.8354$$