الإحصاء والإحتمالات - المحاضرة الثامنة – هندسة

Statistics and probabilities- Lecture 8

Dr Fadi KHALIL

Doctor lecturer in statistics and programing

في هذه المحاضرة سيتم التطرق لـ

- اختبار الفرق بين متوسطي عينتين Testing the equality of means for two samples
- مبدأ تصميم التجارب Experiment design
- تحليل التباين Analysis of variance

11-6- اختبار الفرق بين متوسطي عينتين مستقلتين :

غالبًا ما يهتم المهندسون والعلماء بمقارنة طريقتين أو أسلوبين مختلفين لتحديد ما إذا كان أي منهما ينتج تأثيراً معنوياً على الظاهرة التي يتم ملاحظتها (أو ما يُسمّى الاستجابة). فمثلاً إذ اقترح مطوّر تصاميم في معمل لإنتاج الطلاء تركيبة جديدة بهدف تقليل زمن التجفيف، بحيث تمثل الظاهرة الملاحظة (أو الاستجابة) هنا هي وقت تجفيف. الغرض من الدراسة هو تحديد ما إذا كانت التركيبة الجديدة للطلاء لها تأثير معنوي في تقليل وقت التجفيف. في هذه الحالة، قام مطور المنتج (أو القائم على التجربة) باختيار 10 عينات اختبار بشكل عشوائي لتركيبة معينة و 10 عينات اختبار للتركيبة الأخرى. تم تنفيذ الأعمال على عينات الاختبار بترتيب عشوائي حتى تم تطبيق جميع العينات العشرين. وبالتالي إذا لوحظ فرق ذو دلالة إحصائية في التجربة عشوائية، يمكن أن يكون المجرب واثقًا في الاستنتاج بأن الاختلاف في معالجات تركيبة الطلاء أدّى إلى اختلاف في أوقات التجفيف.

هذه الحالة تسمّى الاستدلال الإحصائي على الفرق بين متوسطي مجتمعين عن طريق دراسة الفرق بين متوسطي عينتين مسحوبتين من هذين المجتمعين ومن ثمّ اختبار فيما وجود فرق بينهما. ففي المثال السابق تمّ الاستدلال عن تأثر طريقة معالجة تركيبة الطلاء في وقت التجفيف وذلك عن طريق اختيار عينتين عشوائيين ودراسة معنوية الفرق بينهما في وقت التجفيف.

بفرض أن متوسطي مجتمعين μ_1، μ_2، يمكن تقديرهما عن طريق سحب عينتين بحجم n_1، n_2 من كل مجتمع على حدا $\bar{x}_1$، و $\bar{x}_2$ ، بالتالي، فإنّ تقدير الفرق بين متوسطي المجتمعين يمكن تقديره $\mu_1 - \mu_2$ بالفرق بين متوسطي العينتين $\bar{x}_1 - \bar{x}_2$

بناءً عليه، ولاختبار الفرق[1] بين متوسطي المجتمعين μ_1، μ_2 يمكن صياغة الفرضية العدم الآتية:

$$H_0 = \mu_1 - \mu_2 = \Delta_0$$

والفرضية البديلة :

$$H_1 = \mu_1 - \mu_2 \neq \Delta_0$$

بعد ذلك يمكن استخدام صيغة مؤشر الاختبار السابقة لتعميمها على حالة الفرق بين المتوسطين:

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

[1] من الجدير للذكر أنّه يمكن اختبار ان يكون الفرق بين متوسطي المجتمعين يساوي الصفر أي يتم اختبار جوهرية أو معنوية الفرق بين متوسطي المجتمعين وتصبح قيمة Δ_0 مساوية للصفر ، $H_0 = \mu_1 - \mu_2 = 0 \ or \ \mu_1 = \mu_2$

بحيث يتم استبدال μ_0 بـ $(\mu_1-\mu_2)$، و قيمة مؤشر العينة $\bar{x}$ بـ $\bar{x}_1-\bar{x}_2$ ، و الخطأ المعياري للمعاينة بـ $\frac{\sigma}{\sqrt{n}}$ بـ $\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$ ، وذلك على افتراض أنّ العينتين مستقلتين بالتالي يصبح مؤشر الاختبار المستخدم لاختبار الفرق بين متوسطي مجتمعين:

$$Z_c=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

بحيث يخضع هذا المؤشر للتوزيع الطبيعي المعياري $N(0,1)$.

وكما تمّ عرضه في اختبار الفرضيات لمتوسط عينة واحدة *one sample test* فمن الممكن أن يكون الاختبار ثنائي الجانب أو أحادي الجانب والجدول الآتي يلخص هذه الحالات مع القيمة المعيارية لاتخاذ القرار:

القيمة المعيارية *critical values*	الفرضيات البديلة *alternative*	الاختبار
$Z_{\frac{\alpha}{2}}$	$H_1=\mu_1-\mu_2\neq\Delta_0$	ثنائي الجانب *two sided*
Z_α	$H_1=\mu_1-\mu_2>\Delta_0$	أحادي الجانب *one sided*
$-Z_\alpha$	$H_1=\mu_1-\mu_2<\Delta_0$	أحادي الجانب *one sided*

- يتم رفض فرضية العدم عموماً إذا كانت قيمة مؤشر الاختبار المحسوب Z_c أكبر من القيمة المطلقة للقيمة المعيارية.

مثال :

بالعود للمثال الأول المتعلق بتطوير تركيبة الطلاء لتحسين وقت التجفيف، حيث تمّ اختبار تركيبتين، الأولى وهي التركيبة المعيارية،أمّا الثانية فهي التركيبة الجديدة والتي يجب أن تخفّض وقت التجفيف. من المعروف من خلال التجربة أنّ الانحراف المعياري لتجفيف الخلطة هو 8 ساعات مهما كان نوعها. فإذا تمّ اختيار 10 عينات من التركيبة الاولى المعيارية و 10 عينات من التركيبة الجديدة، وكان متوسط وقت التجفيف للعينات من التركيبة الأولى هو 36 ساعة بينما كان متوسط وقت التجفيف للعينات من التركيبة الجديدة هو 27 ساعة، فما هي النتيجة التي يمكن استخلاصها من قبل مطوّر التركيبة الجديدة، وذلك عند مستوى دلالة 0.05؟

الحل :

لمعرفة أثر تركيبة الخلطة الجديدة في تحسين وقت التجفيف يجب مقارنة وقت التجفيف عن استخدام كلا التركيبتين المعيارية والجديدة، لذلك تمّ سحب عينتين ومقارنة وقت التجفيف في كلا العينتين.

بفرض أنّ μ_1 متوسط وقت التجفيف عند استخدام التركيبة المعيارية، وأنّ μ_2 متوسط وقت التجفيف عن استخدام التركيبة الجديدة، وفي حال التحسن نتيجة لاستخدام التركيبة الجديدة يجب أن يصبح $\mu_1-\mu_2>0$،

بالتالي يمكن صياغة الفرضية العدم والبديلة كالآتي :

$$H_0 = \mu_1 - \mu_2 = 0$$

$$H_1 = \mu_1 - \mu_2 > 0$$

وبحساب مؤشر الاختبار تحت شرط ففرضية العدم :

$$Z_c = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(36-27)-(0)}{\sqrt{\frac{(8)^2}{10} + \frac{(8)^2}{10}}} = 2.52$$

عند مستوى دلالة 0.05، $Z_{0.05}$ تساوي 1.96 وبالتالي $Z_c > Z_{0.05}$ أي يتم رفض الفرضية العدم والاستنتاج بأنّ التركيبة الجديدة ساعدت بشكل جوهري في تحسين وقت التجفيف.

12 – مفهوم تصميم التجارب *Experiment*

التجارب هي جزء طبيعي من عملية صنع القرار الهندسي والعلمي. لنفترض ، على سبيل المثال ، أن مهندس ميكاترونيكس يحقق في تأثيرات طرق المعالجة المختلفة على متوسط استهلاك الوقود في محركات الدرجات النارية. ستتألف التجربة من تكوين عدة عينات اختبار باستخدام كل من طرق المعالجة المقترحة ثم اختبار استهلاك الوقود. يمكن استخدام البيانات من هذه التجربة لتحديد طريقة المعالجة التي يجب استخدامها لتوفير أقصى متوسط لقوة الانضغاط.

إذا كانت هناك طريقتان فقط من طرق المعالجة ذات الأهمية، فيمكن تصميم هذه التجربة وتحليلها باستخدام طرق الفرضية الإحصائية لعينتين كما تمّ تقديمه في المبحث السابق. أي أن المجرب لديه عامل واحد مثير للاهتمام (نوع تركيبة الطلاء) وهناك طريقتان فقط (أي هنا تجربة أحادية العامل *one factor* ومستويين فقط). إذا كان المجرب مهتمًا بتحديد طريقة المعالجة التي تنتج أقصى استهلاك للوقود، فيمكن تحديد عدد العينات المراد اختبارها، وي ثمّ استخدام اختبار Z أو t لعينتين *two samples* وذلك لتحديد ما إذا كانت الطريقتين مختلفتان . ولكن العديد من التجارب أحادية العامل (*one factor*) لا تفي بغرض الدراسات التي تتضمن أكثر من مستويين من العامل.

على سبيل المثال ، قد يرغب المهندس في التحقق من خمس طرق معالجة مختلفة. وفيما يأتي سيتم توضيح كيف يمكن استخدام مفهوم تحليل التباين (*Analysis of Variance, ANOVA*) لمقارنة الطرق عندما يكون هناك أكثر من مستويين لعامل واحد.

1-12- تصميم التجارب الهندسية *Designing Engineering Experiments*:

تعد تقنيات التصميم التجريبي القائمة على الإحصاء مفيدة بشكل خاص في عالم الهندسة لحل العديد من المشكلات المهمة:، واكتشاف ظواهر أساسية جديدة يمكن أن تؤدي إلى تسويق وتطوير تكنولوجيا ومنتجات جديدة. على سبيل المثال ، في حال تطوير عملية ما *process* جديدة، يمكن تحديد المتغيرات *variables* التي يمكن التحكم فيها والتي تؤثر في هذه العملية، مثل درجة الحرارة والضغط ومعدل التغذية. وبالتالي باستخدام تصميم التجارب *design expirement*، يمكن للمهندسين المختصين تحديد أي مجموعة من هذه المتغيرات لها التأثير الأكبر على أداء هذه العملية. حيث يمكن أن تؤدي نتائج هذه التجربة إلى تحسين العائد، تخفيف التقلبات *variability*، تقليص وقت التصميم والتطوير، وتقليص تكلفة الإنتاج.

في هذا السياق ، تعد طرق التصميم التجريبية *expirement design* مفيدة أيضًا في أنشطة التصميم الهندسي التي يتم خلالها تطوير المنتجات الجديدة وتحسين المنتجات الحالية. حيث تتضمن بعض التجارب المصممة إحصائياً في في الإطار الهندسي:

- تقييم ومقارنة مكونات التصميم الأساسية
- اختيار معايير التصميم التي تؤثر في أداء المنتج، والتي تؤدي في عمل التصميم بشكل جيّد.

عادة ما يتسم تصميم التجارب بصفة التتابع. أي أن التجربة الأولى مع نظام معقد (مثلاً عملية تصنيع) تحتوي على العديد من المتغيرات التي يمكن التحكم فيها، وغالبًا ما تكون تجربة فحص مبدئية مصممة لتحديد تلك المتغيرات الأكثر أهمية. ومن ثمّ تُستخدم التجارب اللاحقة لتنقيح هذه المعلومات وتحديد التعديلات المطلوبة لهذه المتغيرات الحرجة لتحسين العملية. أخيراً، الهدف من المجرب هو التحسين ، أي تحديد مستويات المتغيرات الحرجة التي تؤدي إلى أفضل أداء للعملية.

ولكي لا يبقى شرح مفهوم التجارب في الإطار النظري فقط، ربما أفضل طريقة هو تقديم مثال عملي:

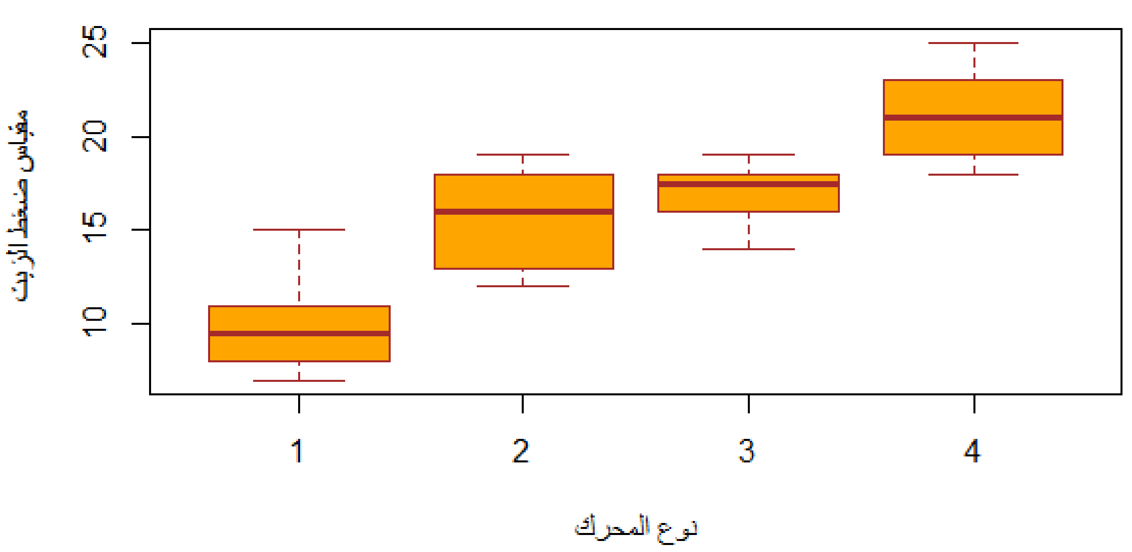
مثال:

بفرض أنّ شركة لتصنيع السيارات قامت بقياس ضغط الزيت في 24 محرك للسيارات موزعة ضمن أربع أنواع أو فئات بحيث تمّ من أجل كل نوع أخذ 6 قياسات بأوقات مختلفة. وكانت المعطيات كالآتي:

نوع المحرك	قياسات الضغط (كيلو باسكال) بالعشرات						الإجمالي	المتوسط
	1	2	3	4	5	6		
1	7	8	15	11	9	10	60	**10**
2	12	17	13	18	19	15	94	**15.67**
3	14	18	19	17	16	18	102	**17**
4	19	25	22	23	18	20	127	**21.17**
المجموع							383	**15.96**

من المفيد مبدئياً عرض الإحصاءات الخاصة بضغط الزيت وذلك بالنسبة لكل فئة من فئات التصدع ، وهذا ما يعبر عنه بأسلوب الرسم البياني *Box plot* حيث يعبر الخط المتواجد داخل الصندوق عن قيمة المتوسط الحسابي لكل فئة، وكذلك تعبر الخطوط الافقية العلوية والسفلية الممتدة على جانبي الصندوق عن أكبر قيمة واصغرقيمة على الترتيب، أمّا حدود الصندوق العلوية والسفلية تعبر عن قيمة الربيع الثالث والأول على الترتيب.

مقاييس ضغط الزيت في محرك السيارة



ومن الواضح في الرسم البياني أنّ الفئة الأخيرة المقابلة للنوع 4 تشكل أعلى نسبة ضغط زيت حيث يتوضع الصندوق المرافق لها في أعلى مكان مقارنة بالصناديق الأخرى. على أيّة حال هناك الكثير من المعلومات يمكن استخراجها من الرسم البياني، ولكن السؤال الأهم يتعلق حول العلاقة بين نوع المحرك مع ضغط الزيت، بمعنى آخر هل تتوزع قيم ضغط زيت المحرك بشكل عشوائي وفقاً لنوع المحرك، أم هناك توزيع لقيم ضغط االزيت يتوافق مع نوع المحرك.

ملاحظة: الحالة هنا تشبه المبحث السابق من ناحية اختبار وجود فرق في وقت التجفف تبعاً لتركيبة الطلاء، حيث تم حينها استخدام اختبار *Z* لاختبار الفرق بين عينتين مستقلتين، نظراص لأنّ العامل *factor* هنا يتضمن مستويين أو تركيبتين فقط، أمّا في هذا المبحث فالعامل يتضمن أكثر من مستوى (4 مستويات للتصدع)، وهنا تقتضي الحالة اتسخدام ما يُسمّى تحليل التباين *Analysis of variance ANOVA* لاختبار معنوية أو جوهرية اختلاف قيم ضغط الزيت وفقاً لنوع المحرك. ونظراً لوجود عامل واحد فقط يسمى تحليل التباين ***One way ANOVA***.

قبل البدء بتحليل التباين لا بدّ من الإشارة إلى ضرورة توفر شرطين وهما:

- أن تكون العينات مستقلة عن بعضها

- أن يكون التوزيع الإحتمالي لقيم المتغير التابع متجانس بين فئات المتغير المستقل.

2-12- تحليل التباين *Analysis of Variance ANOVA*:

بفرض وجود a مستوى لعامل ما وحيد تم تحديدها مسبقاً من قبل المجرب ضمن تجربة معينة وكان هدف هذه التجربة هو مقارنة تأثير مستويات هذا العامل على متغير آخر عشوائي يُسمى *response* . حيث تمّ مشاهدة n قيمة لكل مستوى من مستويات العامل المدروس وذلك بترتيب عشوائي *completely randomized design (CRD)*. وليكن y_{ij} المشاهدة j، للمستوى i من مستويات العامل. ويمكن إظهار ذلك في الجدول الآتي:

مستويات العامل	المشاهدات				الإجمالي	المتوسط
	1	2	$\cdots$	n		
1	y_{11}	y_{12}	$\cdots$	y_{1n}	$y_{1.}$	$\bar{y}_{1.}$
2	y_{21}	y_{22}	$\cdots$	y_{2n}	$y_{2.}$	$\bar{y}_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
α	$y_{\alpha 1}$	$y_{\alpha 2}$	$\cdots$	$y_{\alpha n}$	$y_{\alpha .}$	$\bar{y}_{\square .}$
					$y_{..}$	$\bar{y}_{..}$

من توزيع قيم المتغير y_{ij} في الجدول السابق يمكن القول انّ كل قيمة من قيمة المتغير هي عبارة عن ثلاثة أنواع من العوامل، **النوع الأول** هي عوامل مشتركة بين جميع مستويات العامل المدروس ويرمز له بالرمز μ، **النوع الثاني** خاص بكل مستوى من مستويات العامل المدروس ويرمز له بالرمز τ_i ، أمّا **النوع الثالث** فهو يعبر عن العوامل العشوائية التي لا يمكن حصرها والتي تؤثر في قيم y_{ij} ويمز لها بالرمز ε_{ij}. بمعنى آخر يمكن صياغة المعادلة الآتية:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

بحيث أنّ $i = 1,2, \ldots . a$ و $j = 1,2, \ldots . n$

ويمكن كتابة المعادلة السابقة بشكل آخر كما يأتي:

$$y_{ij} = \mu_i + \varepsilon_{ij}$$

حيث : $\mu_i = \mu + \tau_i$

إذا الهدف من هذه القياسات هو دراسة تأثير مستويات العامل المدروس المحدد مسبقاً من قبل المجرب على المتغير العشوائي y_{ij} وهذا ما يسمى بـ *fixed effect model*.

وبما إنّ العامل من النوع الثاني τ_i خاص بكل مستوى من مستويات العامل المدروس ، فإنّ السؤال المطروح هنا أنّه لو كانت قيم τ_i متساوية بين جميع مستويات a، أي لو كانت قيم μ_i متساوية بين جميع المتسويات، فهل يبقى هناك تأثير لمستويات هذا العامل على قيم y_{ij}؟ الجواب هو قطعا لا .....

إذا المشكلة هنا هي اختبار للفرضيات بمعنى آخر هو اختبار الفرضية العدم الآتية:

$$H_0: \mu_1 = \mu_2 = \cdots \mu_a$$

بينما الفرضية البديلة هي :

$$H_1: \mu_1 \neq \mu_2 \neq \cdots \mu_a$$

وللتذكير هنا أن هذا الاختبار يشبه اختبار تساوي متوسطي عينتين والذي تمّ تقديمه في المبحث السابق، ولكن هنا يتم اختبار التساوي بين أكثر من متوسطين.

لبناء مؤشر الاختبار سيتم اتباع الخطوات الآتية :

أولاً : حساب التباين بين المستويات، *Between*، MS_B :

$$MS_B = \frac{SS_B}{a-1}$$

حيث يعبر a عن عدد مستويات العامل المدروس، و $SS_{Treatments}$ عن مجموع مربعات الفروقات بين المستويات Sum of squares of treatments ويحسب كما يأتي :

$$SS_B = n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2$$

$$\bar{y}_{i.} = \frac{y_{i.}}{N}, \qquad y_{i.} = \sum_{j=1}^{n} y_{ij}$$

$$\bar{y}_{..} = \frac{y_{..}}{N}, \quad y_{..} = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}$$

حيث، N هي عدد المشاهدات الإجمالي ويساوي $n * a$

ثانياً: حساب تباين ضمن المجموعات *Within* ، MS_W

$$MS_W = \frac{SS_W}{a(n-1)}$$

حيث يعبر SS_W عن مجموع مربعات الفرق بين القيم y_{ij} ومتوسط كل مستوى من مستويات العامل المدروس

$$SS_W = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2$$

مع الإشارة إلى انّ مجموع مربعات الفرق بين القيم والمتوسط العام $\bar{y}_{..}$ يسمى بـ SS_T.

$$SS_T = \sum_{i=1}^{\alpha}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$$

وهو يساوي:

$$SS_T = SS_B + SS_W$$

ثالثاً: حساب مؤشر الاختبار:

$$F_0 = \frac{SS_B/a-1}{SS_W/a(n-1)} = \frac{MS_B}{MS_W}$$

حيث يخضع مؤشر الاختبار F_0 لتوزيع فيشر الاحتمال بدرجتي حرية $\alpha - 1$، و $\alpha(n-1)$، وبالتالي لاتخاذ القرار برفض أو عدم رفض فرضية العدم يتم مقارنة القيمة المحسوبة F_0 بالقيمة النظرية أو المعيارية $F_{\alpha,a-1,a(n-1)}$:

فإذا كان: $F_0 > F_{\alpha,a-1,a(n-1)}$ يتم رفض فرضية العدم وبالتالي يوجد أثر لمستويات العامل في قيم المتغير y_{ij} ، امّا في حال كان $F_0 < F_{\alpha,a-1,a(n-1)}$ حينها لا يمكن رفض فرضية العدم وبالتالي مستويات العامل لا تختلف جوهرياً في تأثيرها على قيم المتغير y_{ij}.

يمكن تلخيص ما سبق في الجدول الآتي:

مصدر التباين **variation**	مجموع المربعات ***sum of squares***	درجات الحرية	متوسط المربعات	F_0
بين المجموعات *Between*	SS_B	$a-1$	$MS_B = \frac{SS_B}{a-1}$	$\frac{MS_B}{MS_W}$
ضمن المجموعات *Within*	SS_W	$a(n-1)$	$MS_W = \frac{SS_W}{a(n-1)}$	
الكلي *Total*	SS_T	$an-1$		

وكمثال تطبيقي على تحليل التباين سيتم العودة لبيانات المثال السابق عن دراسة العلاقة بين ضغط االزيت و نوع المحرك في 24 مشاهدة، ولاختبار وجود اختلاف جوهري في توزع قيم ضغط الزيت وذلك تبعاً لنوع المحرك وذلك عند مستوى دلالة 0.01:

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

بينما الفرضية البديلة هي :

$$H_1: \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$$

بحساب SS_T :

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2 = 512.96$$

وكذلك SS_B:

$$SS_B = n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 = 382.79$$

وبالتالي يمكن حساب مجموع مربعات الفروقات ضمن *within* المجموعات SS_W:

$$SS_W = SS_T - SS_B = 512.96 - 382.79 = 130.17$$

بالتالي يمكن بناء جدول تحليل التباين كما يأتي:

مصدر التباين ***variation***	مجموع المربعات ***sum of squares***	درجات الحرية	متوسط المربعات	F_0
بين المجموعات *Between*	382.79	$4 - 1 = 3$	$MS_B = \frac{382.79}{3} = 127.60$	$\frac{\mathbf{127.6}}{\mathbf{6.51}} = \mathbf{19.6}$
ضمن المجموعات *Within*	130.17	$4(6-1) = 20$	$MS_W = \frac{130.17}{20} = 6.51$	
الكلي *Total*	512.96	$4 * 6 - 1 = 23$		

بمقارنة القيمة المحسوبة $F_0 = \mathbf{19.6}$ مع القيمة الجدولية عن مستوى دلالة 0.01 $= F_{\alpha,a-1,a(n-1)}$ $F_{0.01,3,20} = 4.94$ يلاحظ أنّ $F_0 > F_{0.01,3,20}$ وبالتالي يمكن رفض فرضية العدم والتفسير بانّ هناك اتجاه قوي للقول بأنّ ضغط الزيت يختلف اختلافاص جوهريا بين أنواع المحركات المختلفة.

مثال:

قامت مديرية الخدمات بتنفيذ دراسة لإقامة الأحواض المائية، بحيث تمّ دراسة معدل هطول الأمطار في ثلاث مقاطعات

المقاطعة *governate*	معدل هطول الامطار				
	1	2	3	4	5
1	120	151	126	146	160
2	144	135	132	142	125
3	152	172	150	138	256

المطلوب اختبار فيما إذا كان هناك فرق جوهري في متوسط هطول الامطار في المقاطعات الثلاث وذلك عند مستوى دلالة 0.05، إذا علمت أنّ التباين بين المجموعات $SS_B = 863.33$، والتباين خلال المجموعات $SS_W = 1991.57$ ، وأنّ القيمة المعيارية لتوزيع فيشر المرافقة $F_{\alpha,a-1,a(n-1)} = F_{0.05,2,12} = 3.88$.

الحل :

من الواضح هنا أنّ المطلوب هو اختبار وجود فرق جوهري في معدل هطول الأمطار بين فرع وآخر

أي يمكن صياغة فرضية العدم كما يأتي :

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

بينما الفرضية البديلة هي :

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

يمكن اختبار هذه الفرضية عن طريق تحليل التباين بمعامل واحد *one way ANOVA* حيث العامل *factor* يمثل هنا المنطقة الجغرافية وهو يملك ثلاث مستويات أي أنّ a يساوي 3. بينما عدد المشاهدات ضمن كل مستوى من مستويات العامل هو 5 أي $n = 5$، وهذا ما يفضي إلى ان عدد المشاهدات الإجمالي هو $N = n * a = 3 * 5 = 15$.

الخطوة الثانية تتمثل بحساب مربعات الفروق بين المجموعات SS_B، ومربعات الفروق ضمن او خلال المجموعات SS_W.

فإذا كان SS_B يساوي 863.3335 و SS_W يساوي 1991.5785، فإنّ التباين الكلي

$$SS_T = 2854.912$$

حينها يتم الانتقال للخطوة الثالثة وهي حساب مؤشر الاختبار:

$$F_0 = \frac{SS_B/a-1}{SS_W/a(n-1)} = \frac{MS_B}{MS_W}$$

بإنشاء الجدول الآتي:

مصدر التباين **variation**	مجموع المربعات **sum of squares**	درجات الحرية	متوسط المربعات	$\boldsymbol{F_0}$
بين المجموعات *Between*	$SS_B = 863.3335$	$3-1=2$	$MS_B = \frac{863.3335}{2}$	$\boldsymbol{\frac{MS_B}{MS_W} = 2.6}$
ضمن المجموعات *Within*	$SS_W = 1991.5785$	$3(5-1) = 12$	$MS_W = \frac{1991.5785}{12}$	
الكلي *Total*	$SS_T = 2854.912$	$3*5-1 = 14$		

أي أنّ القيمة المحسوبة لمؤشر الاختبار هي 2.6 مستوى دلالة 0.05 هو $F_{\alpha,a-1,a(n-1)} = F_{0.05,2,12} =$ 3.88 يلاحظ أنّ $F_0 < F_{0.05,2,12}$ وبالتالي لا يمكن رفض فرضية العدم والتفسير بانّ متوسط هطول الامطار لا يختلف بين منطقة جغرافية وأخرى.