الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي والبحث العلمي

جامعة الاخوة منتوري قسنطينة.01.

مـــــادة

الاحصــــــــــــــــا

مستوى: سنة ثالثة ليسانس.

تخصص: بيولوجيا وفيسيولوجيا النبات.

مسؤول المقياس: الاستاد عيسى جروني.

مايل: aissa.djerouni@umc.edu.dz

قسم: بيولوجيا وايكولوجيا النبات.

كلية: علوم الطبيعة والحياة.

الفهرس

الفصل الأول:

- مفهوم علم الاحصاء.
- المقدمة.
- الهدف من دراسة مادة الاحصاء.
- تعريف علم الاحصاء.
- الطريقة الاحصائية في البحث العلمــي.
- عرض النتائج بالطريقة البسيطة.
- المدرج التكراري.
- المضلع التكراري.
- المنحنى التكراري.

الفصل الثاني:

- اختبار الفرق بين متوسطي مجتمعين.
- تدكير بفرضيات العدم والبديلة.
- المتوسط الحسابي.
- تباين العينة.
- الخطأ المعياري.
- التباين التجميعي.
- اختبار الفرضية t^* المحسوبة.
- القرار النهائي لتجربة.
- مثـــال.

الفصل الثــالث:

- اختبار الفرق بين متوسطين في حالة البيانات المتزاوجة.
- الوسط الحسابي.
- الخطأ المعياري.
- الوسط الحسابي للفروق.
- الانحراف المعياري.
- القرار النهائي لتجربة.
- مثــال.

الفصل الرابع:

- مقدمة.
- الفرضيات المطلوب اختباره.
- تحليل التباين الاحادي.
- معامل التصحيح.
- حساب مجموع المربعات الكلية.
- مجموع مربعات التباين بين المجموعات.
- مجموع المربعات داخل المجموعات.
- متوسط المربعات بين المجموعات.
- متوسط المربعات داخل المجموعات
- حساب F^*.
- اخد قرار بالاختلاف المعنوي من عدمه في التجربة.
- مثــال.

مفهـــــــــــوم علم الإحصــــــــاء

المقدمة:

اليوم، من اهم العلوم التي تتوقف عليها التنمية السياسية والاقتصادية والثقافية، وللإحصاء حصة اساسية من عمل الدول والمؤسسات والمنظمات السياسية والاقتصادية والاجتماعية، عالميا ودوليا ومحليا، وكثيرا ما يرتهن مصير مشاريع او قرارات كبرى بالنتائج التي يقدمها الاحصاء في مجال معين.
وبصورة عامة، فان افتقاد الجهد الاحصائي، في مجال من المجالات، يمنع من التأكد وتحصيل الضمان في استجابة اي مشروع للواقع، كما يحول دون تحديد مدى نجاحها و اخفاقها، ويجعل في الاقدام عليها شيئا من المخاطرة.

الهدف من دراسة مادة الاحصاء:

تهدف دراسة مادة الاحصاء الى:

* تعريف الطالب بطبيعة علم الاحصاء وأهميتها ومجالات استخدامها.
* تعريف الطالب بأساليب جمع البيانات وطرق اخذ العينات ومن ثم طرق عرضها بالطرق البيانية.
* تعريف الطالب باتباع الاساليب الاحصائية لغرض الوصول الى النتائج الدقيقة بأقصر طريقة واقل كلفة.
* للإحصاء دور مهم واساسي في التخطيط للمشاريع سواء اكانت مشاريع فردية ابحاث ام مشاريع تخص المجتمع استخدام الاساليب الاحصائية تساعد الطالب مستقبلا على اتخاذ قراراته سواء اكانت في تجربة علمية لزيادة انتاج محصول او تأثير لمكافحة مرض المحاصيل.
* توضيح للطالب كيف ان الاحصاء تغير من صورته القديمة الموجودة في اذهان الناس على انها علم العد وجمع البيانات وعرضها الى اعتباره الان علم يعمل على استخدام الاسلوب العلمي في طرق جمع البيانات وتبويبها وتحليلها وعرضها لاستخلاص النتائج واتخاذ القرار المناسب على ضوء ذلك.

من المتعارف عليه أن العلماء في مختلف المجالات يستعملون الطريقة العلمية في دراساتهم وأبحاثهم. فما هي هذه الطريقة العملية؟ وللإجابة على هذه السؤال نجد انه من الصعب اعطاء تعريف عام وموحد يصف الطريقة العلمية. والسبب في ذلك ان العلماء المختلفين يستعملون كل الطرق والوسائل المختلفة المتاحة لهم والتي يعتقدون انها سوف تساعدهم في الوصول الى غرضهم ألا وهو فهم وتفسير الظواهر الطبيعية المختلفة التي يقومون بدراستها. ومن ناحية اخرى فإن الوسائل المختلفة التي يستعملها الباحثون في دراستهم تشترك في عدة صفات عامة يمكن تلخيصها في الاتي:

1- عندما يصادف الباحث ظاهرة طبيعية لم يصادفها من قبل، أي أن هذه الظاهرة لم تكن معروفة له من قبل فإنه يحاول أن يفهمها وأن يفسرها. مثال ذلك عندما بدأ المشتغلون في مجال مقاومة الافات الزراعية يلاحظون أن مبيدا حشريا معينا يصبح بعد فترة من استعماله عديم التأثير على الحشرات التي كان يقضي عليها قبلا....فإن هذا يحتاج الى تفسير معقول ومقبول.
2- بعد التأكد من حقيقة حدوث هذه الظاهرة الطبيعية وأنها ليست ظاهرة عرضية أو فتية فإن الخطوة الاولى في تفهم هذه الظاهرة تكون باستعراض كل الحقائق والنضريات والفروض المختلفة التي لها علاقة بهذه الظاهرة الجديدة.
3- بعد ذلك يقوم الباحث بوضع نضرية فرضية Null hypothesis H_0 لتفسير هذه الظاهرة الجديدة وارجاعها الى اسبابها. وتستند هذه النضرية على حقائق معروفة من قبل وذلك لها علاقة بالضاهرة الجديدة.
4- يجب أن تكون النضرية الفرضية H_0 الموضوعة قابلة للإثبات التجريبي في المعمل وإلا فإنها تفقد قيمتها العلمية فتكون غير مقبولة اصلا.
5- يقوم الباحث بفرض عدة فروض على اساس ان النضرية الفرضية صحيحة ثم يقوم بالتحقق من صحة هذه الفروض تجريبيا فإذا ثبت صحة هذه الفروض زاد عدد الفروض المثبتة تجريبيا كلما زاد التأكد من صحة هذه النضرية الفرضية.

الخطوة الاخيرة في هذه السلسة تكون اقتناع الباحث بصحة نظريته الفرضية. وعليه بعد ذلك اقناع باقي زملاءه الذين يعملون في مجال تخصصه. وعادة تكون هذه أصعب الخطوت..

طبيعة علم الاحصاء.

كلمة) الاحصاء (في الماضي كانت تهدف الى العد والحصر حتى سمي الاحصاء بعلم العد اما الاحصاء الأن فقد تطور كثيرا وخاصة في القرن العشرين واصبح علما(The Science of canting) . مستقلا له اهميته كوسيلة واداة في البحث العلمي لجميع العلوم

تعريف علم الحصاء3-1 :

هناك تعاريف عديدة للإحصاء اختلفت وتباينت من حيث المضمون والشمول باختلاف مراحل تطوير هذا العلم

كما قد عرف بأنه الطريقة العلمية التي تختص بجمع البيانات والحقائق بشكل يسهل عملية تحليلها وتفسيرها ومن ثم استخلاص النتائج واتخاذ القرار على ضوء ذلك ، وهناك من عرفه بانه العلم الذي

يعمل على استخدام الأسلوب العلمي في طرق جمع البيانات وتبويبها وتلخيصها وعرضها وتحليلها بهدف الوصول منها الى استنتاجات وقرارات مناسبة.

.الطريقة الاحصائية في البحث العلمي

استخدام الاسلوب الاحصائي في البحث العلمي يعني توفير البيانات والمعلومات عن الظاهرة المطلوب دراستها في ذلك البحث وهذا يعني ان امكانية تطبيق الطريقة الاحصائية مرهونا بإمكانية التعبير عن هذه الظاهرة أو تلك تعبيرا كميا) رقميا. (وتمتاز الطريقة الاحصائية بكونها تهيئ اسلوبا موضوعيا محايدا للبحث له قواعده واصوله التي يجب أن يلتزم بها الباحث حتى يتجنب التحيز الشخصي والوقوع في بعض الاخطاء . كما يساعد استخدام الطريقة الإحصائية الى وصول الباحث الى النتائج الدقيقة بأقصر طريق وأقل كلفة.

عرض النتائج بالطريقة البسيطة:

.2العرض البياني للبيانات المبوبة

العرض البياني للبيانات المبوبة :ان الرسوم والاشكال الهندسية ما هي الا تعبير وتوضيح للبيانات بطريقة جذابة وسهلة وفعالة تساعد القارئ على فهم واستيعاب الظاهرة ومقارنتها مع بعضها . ووسائل التمثيل البياني كثيرة ومتنوعة وسنكتفي بشرح العرض البياني للتوزيعات التكرارية فقط . وعادة نخصص المحور الافقي او الاحداثي السيني لتمثيل قيم او فئات المتغير بينما نخصص المحور العمودي او الاحداثي الصادي لتمثل تكرارات هذا المتغير ويجب دائما ان يبدأ تدريج المحور العمودي من الصفر اما تدريج المحور الافقي فقد لا نبدأ بتدريجه من الصفر .ومن اشكال العرض البياني نذكر.

.المدرج التكراري

عبارة عن مستطيلات رأسية تمتد قواعدها على المحور الافقي لتمثل اطوال الفئات بينما ارتفاعاتها تمثل تكرارات الفئات .ولرسم مدرج تكراري نتبع الخطوات الاتية:

1-رسم المحور الافقي والمحور العمودي

2-تدريج المحور الافقي الى اقسام متساوية بحيث يشمل جميع حدود الفئات ويفضل ترك مسافة صغيرة بين نقطة الصفر والحد الادنى للفئة الاولى، ونقسم المحور العمودي الى اقسام متساوية بحيث تشمل اكبر

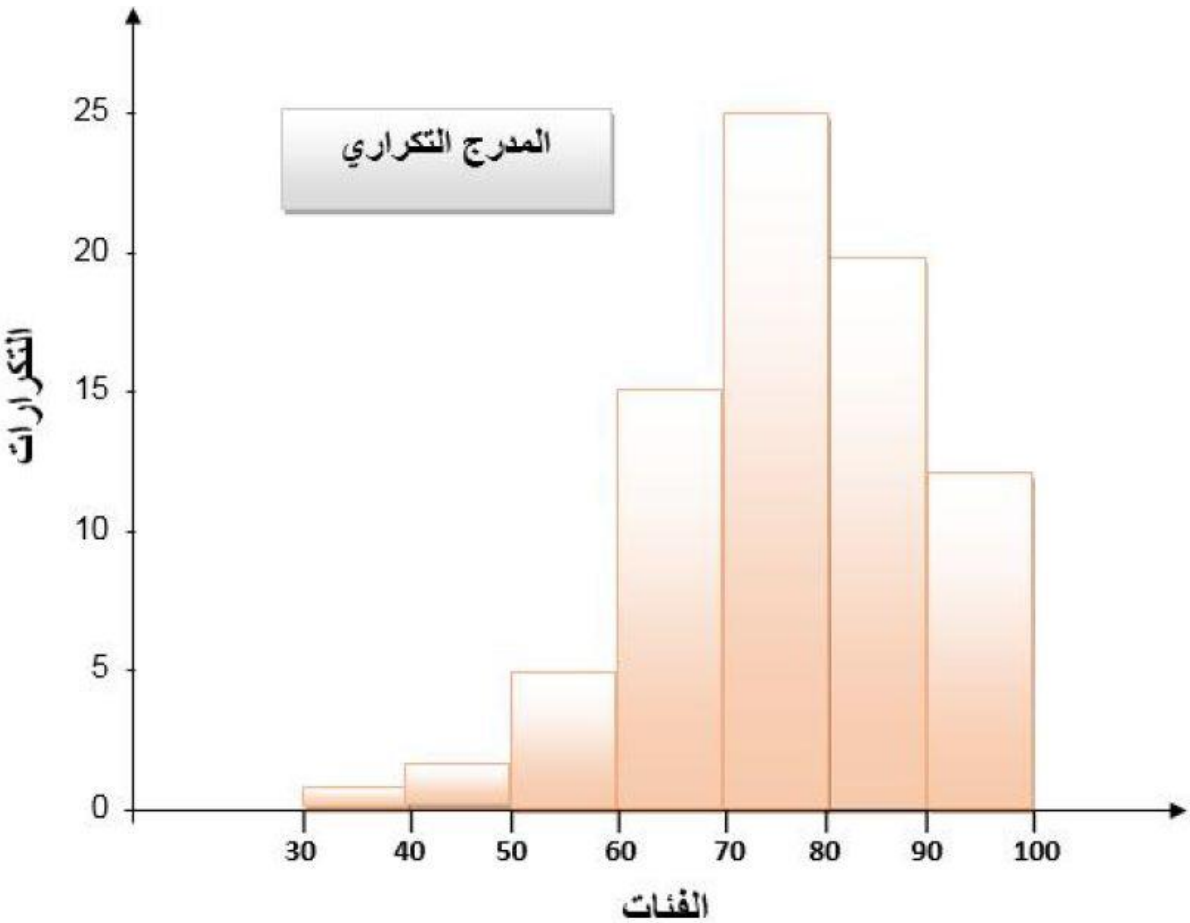
التكرارات

3-يرسم على كل فئة مستطيلا رأسيا تمثل قاعدته طول تلك الفئة وارتفاعه يمثل تكرارها

مثال: (1) الجدول الاتي يمثل توزيع تكراري لأطوال نباتات القطن/ المطلوب : تمثيل التوزيع التكراري

الفئات	30-40	40-50	50-60	60-70	70-80	80-90	90-100	**المجموع**
التكرارات	1	2	5	15	25	20	12	**80**

بمدرج تكراري



.المضلع التكراري:

هو وسيلة اخرى لتمثيل التوزيع التكراري بيانيا ويمكن رسمه بإحدى طريقتين اولهما: اذا كان المدرج التكراري معلوم . ويتم ذلك بتصنيف القواعد العليا لمستطيلات المدرج ثم نصل بين هذه النقط بمستقيمات ورسم فئة قبل الاولى تكرارها صفر وفئة بعد الاخيرة تكرارها صفر وبتصنيف هاتين الفئتين وتوصيل بقية الخط نحصل على ما يسمى بالمضلع التكراري.

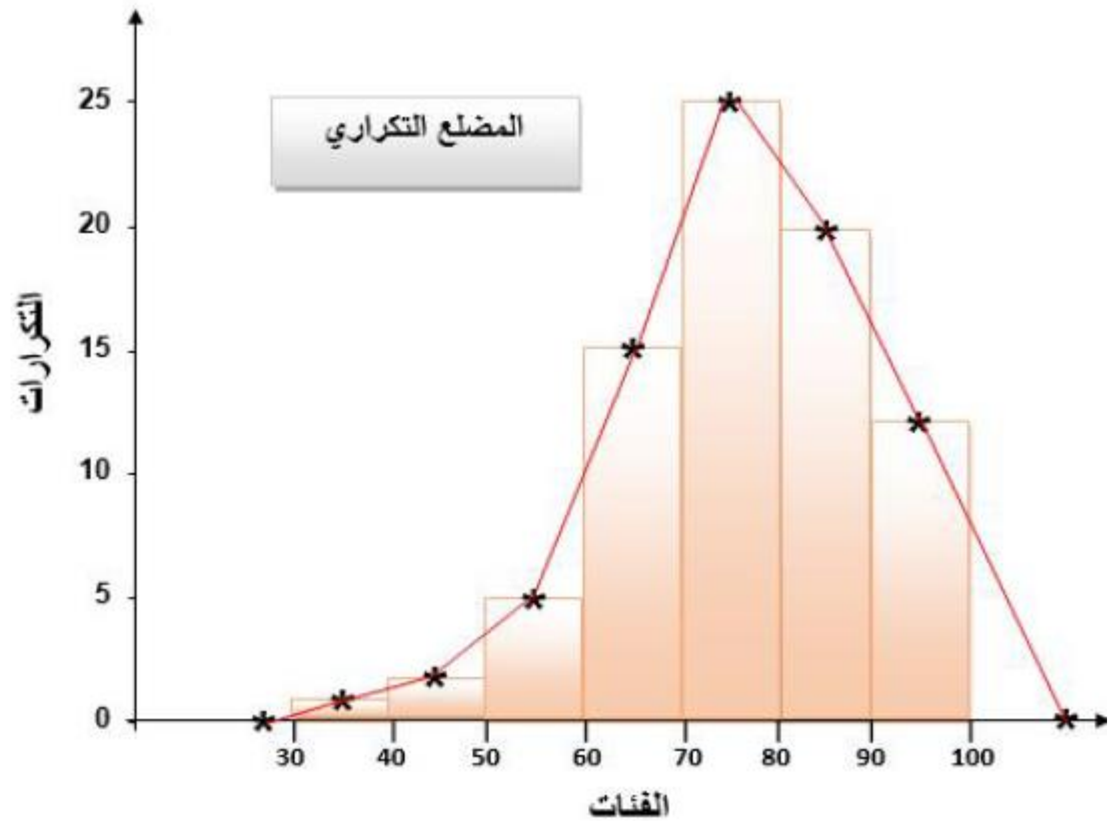
مثال : (3) ارسم ما يلي للبيانات التالية:

1-المضلع التكراري بطريقتين

الفئات	30-40	40-50	50-60	60-70	70-80	80-90	90-100	المجموع
التكرارات	1	2	5	15	25	20	12	**80**

2-المنحني التكراري

الطريقة الأولى:



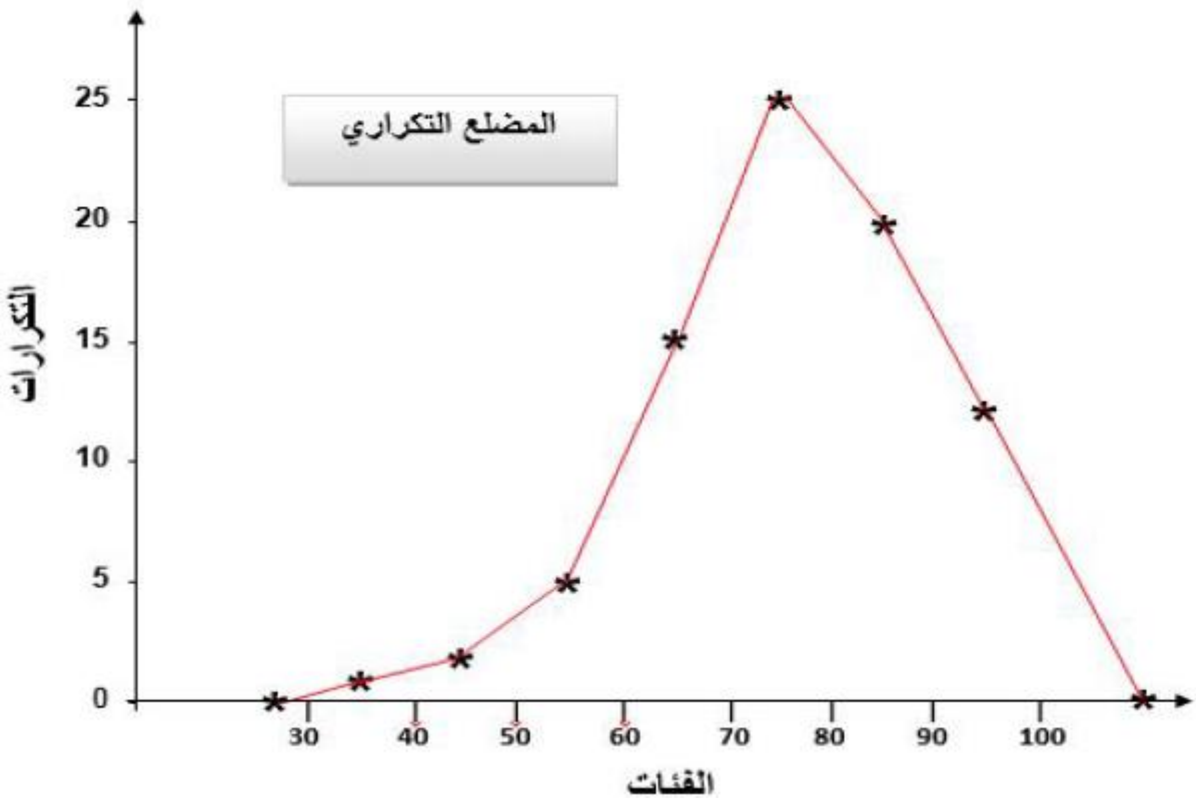
الطريقة الثانية : رسم المضلع التكراري على مراكز الفئات مباشرة دون ضرورة لرسم المدرج التكراري أولا عليه فان المحور الأفقي يمثل مراكز الفئات والمحور العمودي يمثل التكرارات ثم نصل النقاط بعضها ببعض وعليه فان خطوات رسم المضلع التكراري كما يأتي:

1-يجاد مراكز الفئات على المحور الأفقي

2-تحديد النقطة التي تقابل مركز كل فئة على المحور الرأسي

– 3وصل مستقيمات بين النقط التي حددناها بعضها ببعض

وعليه فان المضلع التكراري يكون بالشكل التالي



إختبار الفرق بين متوسطي مجتمعين:

لنفترض أن لدينا مجتمعين مختلفين ونريد أن نختبر ما إذا كان متوسطا هذين المجتمعين متساويين أو أن نضع فترة الثقة للفرق بين متوسطين. ولنرمز لمتوسطي المجتمعين بالرمز μ_1 و μ_2 وللتباين بـالرمز ϵ_1^2 و ϵ_2^2 ويكون الفرق العدم الذي نريد اختباره هو: $H_0: \mu_1 = \mu_2$

ضد أي من الفرضيات البديلة التالية:

$$H_a: \mu_1 \neq \mu_2 \quad , \quad H_a: \mu_1 > \mu_2 \quad , \quad H_a: \mu_1 < \mu'_2$$

وللاستدلال عن هذه فرض العدم نقوم بجمع البيانات تتمثل في عينتن عشوائيتين من المجتمعين تحت الدراسة، وهي عبارة عن قياسات المتغير عشوائين X_1 و X_2 ونرمز لتلك البيانات كمايلي:

عينة من المجتمع 1 :	$X_{11} \quad X_{12} \quad \dots \quad X_{1n_1}$
عينة من المجتمع 2 :	$X_{21} \quad X_{22} \quad \dots \quad X_{2n_2}$

ونرمز للمقاييس الوصفية لكل مجموعة على النحو التالي: حجم العينة: n_1 و n_2.

متوسط حسابي : $\overline{X}_1$, $\overline{X}_2$ حيث $i = 1, 2$ $\quad \overline{X}_i = \frac{\sum_{j=1}^{n_i} X_{ij}}{n_i}$

تباين العينـــة : S_1^2 S_2^2 حيث $\quad S_i^2 = \sum_{j=1}^{n_i} (X_{ij} - \overline{X}_i)^2/(n_i - 1)$

وتعتبر $\underline{X}_1$ و $\underline{X}_2$ تقديرا لكل من μ_1 و μ_2 كما تعتبر S^2_1 و S^2_2 تقديرا لكل من σ_1^2 و σ_2^2.

لإجراء اختبار الفرق بين متوسطي المجتمعين نضع الافتراضات التالية:

أن المجتمعين طبيعيان أي أن $i= 1,2$.

1- أن المجتمعين طبيعيان : أي أن $X_i \sim N(\mu_i , \sigma_i^2)\ ; i = 1, 2$.

2- أن تبايني المجتمعين متساويان أي $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

وإذا ما توفرت هذه الافتراضات فنستخدم الاختبار الإحصائي t التالي :

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_{\overline{x}_1 - \overline{x}_2}}$$

الذي يتبع توزيع t بدرجات الحرية (n_1+n_2-2).

وتعرف بالقيمة $S_{\overline{x}_1 - \overline{x}_2}$ بالخطأ المعياري (Standard error) للفرق بين المتوسطين، وهي تقدير للإنحراف القياسي للفرق بين متوسطين $\sigma_{\overline{x}_1 - \overline{x}_2}$ ويحسب الخطأ المعياري في ظل الافتراض الثاني السابق كتالي:

$$S_{\overline{x}_1 - \overline{x}_2} = S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

حيث

$$S_p^2 = \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{(n_1 + n_2 - 2)}$$

وتسمى S_p^2 بالتباين التجميعي (Pooled Variance) أو التباين المرجح لتباين العينتين S_1^2 و S_2^2 وهي تقدير للتباين σ^2 المعرف في الافتراض الثاني. وتحسب قيمة الاختبار الاحصائي تحت ماورد ذكره في H_0 وما توفر لدينا من مقاييس وصفية للعينتين، بحيث يصبح الاختبار المحسوب هو:

$$t^* = \frac{(\overline{X}_1 - \overline{X}_2)}{S_{\overline{x}_1 - \overline{x}_2}}$$

ثم نقارن قيمة t^* مع قيمة t الجدولية لإصدار القرار حول H_0 بناء على الحالة التي تم تحديدها من الحالات الثلاث للفرض البديل H_a. ويلخص الجدول (2-1) قرارات اختبار مقارنة متوسطي مجتمعين. ونعرف قيمة t الجدولية المستخدمة في الجدول (1.2).

جدول (2-1): قرارات اختبار t المقارنة متوسطي مجتمعين.

فرض العدم $H_0: \mu_1 = \mu_2$	القرار
الفرض البديل : $H_a: \mu_1 \neq \mu_2$	نرفض H_0 إذا $\lvert t^* \rvert > t_{(1-\alpha/2, n_1 + n_2 - 2)}$
$H_a: \mu_1 > \mu_2$	نرفض H_0 إذا $t^* > t_{(1-\alpha, n_1 + n_2 - 2)}$
$H_a: \mu_1 < \mu_2$	نرفض H_0 إذا $t^* < t_{(\alpha, n_1 + n_2 - 2)}$

ويفي هذا الاختبار بالهدف الاول للتجارب العلمية الموضح في الفصل الاول والمتمثل في الاجابة عن السؤال: هل هناك فرق معنوي بين المتوسطات ؟ وأما الهدف الثاني للتجارب العلمية فهو تقدير الفرق ان وجد. أي اذا كانت هناك اختلافات معنوية بين μ_1 و μ_2.

مثال :

في دراسة لمقارنة العناصر المعدنية لنوعين من العصائر، عصير البرتقال وعصير التفاح، أخذت عينتان عشوائيتان من العلب المعروضة في الاسواق لكل منهما، ومن بين القياسات نذكر قياسات كمية الصوديوم وكانت البيانات على النحو الذي بالجدول 2-2.

ونريد اختبار فرض العدم $H_0: \mu_1 = \mu_2$ ضد الفرض البديل $H_a: \mu_1 \neq \mu_2$ عند $\alpha = .05$. فيحسب التباين التجميعي S_p^2 كما يلي :

جدول (2-2) : محتويات عصير البرتقال و التفاح من الصوديوم (mg/100gr)

	التفاح (1)	البرتقال (2)
	4.86	4.72
	5.11	4.81
	5.23	5.22
	5.19	5.67
	5.61	5.52
	5.32	4.96
	5.20	5.35
	4.95	5.34
	4.98	
n_i	9	8
$\overline{X}_i$	5.16	5.20
S_i^2	0.0506	0.115

$$S_p^2 = \frac{(9-1)0.0506 + (8-1)0.115}{9+8-2} = 0.0806$$

والخطأ المعياري للفرق بين المتوسطين هو

$$S_{\overline{x}_1 - \overline{x}_2} = S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= \sqrt{.0806 \left(\frac{1}{9} + \frac{1}{8}\right)} = 0.138$$

وقيمة t المحسوبة هي :

$$t^* = \left(\overline{X}_1 - \overline{X}_2\right) / S_{\overline{X}_1 - \overline{X}_2} = (5.16 - 5.20) / (0.138) = 0.29$$

وعند مستوى المعنوية $\alpha = .05$ تكون قيمة t الجدولية الملائمة للحالة الأولى التي بالجدول (2-1) : $t_{(.975,15)} = 2.131$ وبمقارنة قيمة t^* مع 2.131 لا نرفض H_0 ونقر بأن كمية الصوديوم متساوية في كلا العصيرين بناء على البيانات المتاحة بالجدول (2-2) .

اختبار الفرق بين متوسطين في حالة اليانات المتزاوجة:

غالبا مايحاول الباحث الزيادة في دقة التجربة التي تقارن بين معالجتين عن طريق استخدام الازواج. وهذا المفهوم هو نفس مفهوم القطاعات التي تطرقنا له. وهو عبارة عن إيجاد أزواج يكون أفراد كل زوج منها متماثلين أو متجانسين فيما بنهما، وبالتالي يكون الفرق في الاستجابة راجعا بدرجة كبيرة الى الفرق بين المعالجتين المراد مقارنة تأثيرهما.

تحصل البيانات المتزاوجة في التجارب الطبية عندما يعطى دواء لمجموعة من المرضى وتقاس ظاهرة معينة قبل تعاطي الواداء وبعده وفي هذه الحالة تكون قد استخدمت نفس الوحدات التجريبية للمعالجتين * قبل الدواء، بعد الدواء.

وفرض العدم الذي نريد اختباره هنا هو أن متوسط الفروق يساوي صفر (أو أي قيمة معينة) أي $H_0 : \mu_D = 0$ ضد أيّ من الفرضيات البديلة الثلاث $H_a : \mu_D \neq 0$, $H_a : \mu_D > 0$ و $H_a : \mu_D < 0$. ونفترض في هذا الاختبار أن تكون الفروق موزعة حسب التوزيع الطبيعي ، أي أن $D_i \sim N(\mu_D, \sigma_D^2)$. ويبقى هذا الاختبار صالحاً حتى في حالة عدم تجانس التباين : $\sigma_1^2 \neq \sigma_2^2$.

وتتلخص خطوات الاختبار في حساب كل الفروق بين فردي كل زوج D_i , $i = 1,\ldots,n$ والوسط الحسابي والخطأ المعياري للقيم D_i ، ثم إيجاد قيمة

الاختبار الاحصائي:

$$t = \frac{\overline{D} - \mu_D}{S_{\overline{D}}} \sim t_{(n-1)}$$

حيث $\overline{D}$: الوسط الحسابي : $\overline{D} = \sum_{i=1}^{n} D_i / n$

μ_D : متوسط مجتمع الفروق

$\overline{S}_D$: الخطأ المعياري $S_{\overline{D}} = \frac{S_D}{\sqrt{n}}$, $S_D = \sqrt{\frac{\sum_{i=1}^{n}(D_i - \overline{D})^2}{n-1}}$

n : عدد الأزواج .

ويلخص الجدول (3-2) قرارات اختبار t لمقارنة متوسطين في حالة البيانات المتزاوجة . وترمز t^* إلى قيمة الاختبار المحسوبة تحت الظروف المحددة في فرض العدم ومن نتائج البيانات ، بحيث تصبح $t^* = \overline{D} / S_{\overline{D}}$.
وتحسب فترة الثقة لمتوسط الفروق μ_D بمستوى ثقة $100(1-\alpha)\%$ كالتالي :

$$\overline{D} \pm t_{(1-\alpha/2,\, n-1)} S_{\overline{D}}$$

جدول (3-2) : قرارات اختبار t للفرق بين متوسطين في حالة البيانات المتزاوجة .

فرض العدم $H_0 : \mu_D = 0$	القرار	
القرض البديل $H_a : \mu_D \neq 0$	نرفض H_0 إذا	$\lvert t^* \rvert > t_{(1-\alpha/2,\, n-1)}$
$H_a : \mu_D > 0$	نرفض H_0 إذا	$t^* > t_{(1-\alpha,\, n-1)}$
$H_a : \mu_D < 0$	نرفض H_0 إذا	$t^* < t_{(\alpha,\, n-1)}$

ويوضح المثال التالي استخدام اختبار t لمقارنة متوسطين في حالة البيانات المتزاوجة .

مثال 2:

أجريت تجربة لمقارنة نوعية الاكل وتأثيرهما على نمو العجول خلال شهر من التغذية وسجلت الزيادة في الاوزان لسبعة ازواج من العجول، ويوضح الجدول 4-2 تلك البيانات.

ويريد الباحث اختبار $H_0 : \mu_D = 0$ ضد $H_a : \mu_D > 0$ عند مستوى المعنوية $\alpha = 0.05$.

وفيما يلي خطوات اختبار t :

الوسط الحسابي للفروق هو : $\overline{D} = \frac{\Sigma D_i}{n} = \frac{48.4}{7} = 6.91$

الانحراف المعياري للفروق هو $= \quad S_D = \sqrt{\frac{\Sigma D_i^2 - (\Sigma D_i)^2/n}{n-1}}$

$$= \left(\frac{343.84 - (48.4)^2/7}{7-1} \right)^{1/2} = 1.237$$

جدول (4-2) : زيادة أوزان العجول باستخدام عليقتين

Pair	عليقة 1 X_{1i}	عليقة 2 X_{2i}	الفرق $D_i = X_{2i} - X_{1i}$
1	30.51	36.32	5.81
2	29.37	37.51	8.14
3	28.72	35.47	6.75
4	31.33	38.20	6.87
5	31.56	36.52	4.96
6	29.80	37.22	7.42
7	30.50	38.95	8.45
مجموع	211.79	260.19	48.4

والخطأ المعياري هو : $S_{\overline{D}} = S_D / \sqrt{n} = 1.237 / \sqrt{7} = 0.468$

وقيمة t المحسوبة هي : $t^* = \overline{D} / S_{\overline{D}} = 6.91 / 0.468 = 14.8$

والمعنوية المحسوبة هي : P - value < . 0005

وبمقارنة *t بقيمة t الجدولية ، عند درجات الحرية 6 ومستوى المعنوية α = .05 أي $t_{(1-\alpha, n-1)} = t_{(.95,6)} = 1.943$ ، وعلى ضوء القرار الثاني الموضح بالجدول (2–3) نرفض فرض العدم H_0 . وبالتالي فإن العليقة الثانية أفضل من الأولى من حيث زيادة أوزان العجول .

ويمكن تنفيذ هذا الاختبار عن طريق فترة الثقة . فإذا احتوت فترة الثقة على متوسط مجتمع الفروق المعرّف في فرض العدم أي $\mu_D = 0$ فتقبل H_0 ، وأما إذا لم تحتو على الصفر فنرفض فرض العدم . ونحسب لهذا المثال 95% فترة الثقة للمتوسط μ_D كالتالي :

$$\overline{D} - t_{(1-\alpha/2, n-1)} S_{\overline{D}} \leq \mu_D \leq \overline{D} + t_{(1-\alpha/2, n-1)} S_{\overline{D}}$$

$$6.91 - 1.943(0.468) \leq \mu_D \leq 6.91 + 1.943(0.468)$$

$$6.00 \leq \mu_D \leq 7.82 \text{ kg}$$

وبما أن هذه الفترة لا تحتوي على الصفر فنرفض H_0 ونستنتج أن متوسط الفروق لا يساوي صفراً .

مقارنة تبايني مجتمعين:

تأكد من صحة الافتراض $\sigma_1^2 = \sigma_2^2$ بالنسبة لاختبار t لمقارنة متوسطي مجتمعين، فهناك العديد من الأبحاث التي يكون هدفها الرئيسي هو مقارنة σ_1^2 مع σ_2^2 مثل دراسات جودة البضائع المستهلكة حيث يعتبر التباين من أهم مقاييس الجودة.

والفرضية التي نريد اختبارها هنا هي: $H_0 : \sigma_1^2 = \sigma_2^2$

ضد الفرضيات البديلة الثلاث الممكنة وهي $H_a : \sigma_1^2 \neq \sigma_2^2$, $H_a : \sigma_1^2 > \sigma_2^2$ وأخيرا $H_a : \sigma_1^2 < \sigma_2^2$. فلو أخذنا عينة عشوائية $X_1, X_2, ..., X_{n_1}$ من توزيع طبيعي $N(\mu_1, \sigma_1^2)$ وعينة عشوائية أخرى $Y_1, Y_2, ..., Y_{n_2}$ من توزيع طبيعي $N(\mu_2, \sigma_2^2)$ وحسبنا من كل عينة التباين S_i^2 فإن القيمة

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

تكون موزعة حسب توزيع F بدرجات حرية $n_1 - 1$ و $n_2 - 1$. وبالتعويض داخل معادلة F عن S_i^2 المحسوبة من العينتين وإدخال ما ورد في فرض العدم من أن $\sigma_1^2 = \sigma_2^2$ ، يصبح الاختبار الإحصائي كالتالي :

$$F^* = S_1^2 / S_2^2$$

وتقارن قيم *F مع قيم F الجدولية على النحو الموضح بالجدول (2-5) . ونعرف $F^A_{\upsilon_1, \upsilon_2}$ بأنها قيمة F الجدولية بدرجات حرية υ_1 و υ_2 والتي تترك على يسارها مساحة A من التوزيع . انظر الجدول A-4 في الملحق ولا يحتوي هذا الجدول على قيم F التي تترك على يسارها مساحات صغيرة أي أقل من 0.50 وبإمكان القارىء استنتاجها من الجداول بواسطة المعادلة التالية : إذا كانت $\alpha <$ 0.50 فإن

$$\alpha < 0.50 \;:\; F^{\alpha}_{(\upsilon_1, \upsilon_2)} = 1 / F^{1-\alpha}_{(\upsilon_2, \upsilon_1)}$$

جدول (2-5) : قرارات اختبار F لمقارنة تبايني مجتمعين .

فرض العدم : $H_0 : \sigma_1^2 = \sigma_2^2$	القرار
الفرض البديل : $H_a : \sigma_1^2 \neq \sigma_2^2$	نرفض H_0 إذا $F^* > F^{1-\alpha/2}_{(n_1-1, n_2-1)}$ أو $F^* < F^{\alpha/2}_{(n_1-1, n_2-1)}$
$H_a : \sigma_1^2 > \sigma_2^2$	نرفض H_0 إذا $F^* > F^{1-\alpha}_{(n_1-1, n_2-1)}$
$H_a : \sigma_1^2 < \sigma_2^2$	نرفض H_0 إذا $F^* < F^{\alpha}_{(n_1-1, n_2-1)}$

ويعتبر اختبار F لمقارنة تبايني مجتمعين شبيهاً باختبار Bartlett لتجانس التباين الموضح في الفصل الثالث وذلك لأن كليهما يساعدنا على التأكد من صحة الافتراض الأساسي للطريقتين الإحصائيتين : اختبار t لمقارنة متوسطي مجتمعين وتحليل التباين .

وتحسب فترة الثقة للنسبة σ_1^2 / σ_2^2 كما يلي :

$$\frac{S_1^2}{S_2^2} \frac{1}{F^{1-\alpha/2}_{(n_1-1,\, n_2-1)}} \leq \sigma_1^2 / \sigma_2^2 \leq \frac{S_1^2}{S_2^2} \frac{1}{F^{\alpha/2}_{(n_1-1,\, n_2-1)}}$$

وكمثال لتطبيق اختبار F نستخدم بيانات المثال 2-1 للتأكد من سلامة اختبار وتلخص مقاييس المثال فيما يلي:

$$n_1 = 9 \qquad S_1^2 = 0.0506$$
$$n_2 = 8 \qquad S_2^2 = 0.1150$$

ولاختبار فرض العدم $H_0 = \sigma_1^2 = \sigma_2^2$ ضد الفرض البديل $H_a = \sigma_1^2 \neq \sigma_2^2$ عند مستوى المعنوية $\alpha = .05$ نستخدم قيمة F المحسوبة:

$$F^* = \frac{S_1^2}{S_2^2} = \frac{0.0506}{0.1150} = 0.44$$

ونقارنها بقيم F الجدولية التالية :

$$F^{1-\alpha/2}_{(n_1-1,\, n_2-1)} = F^{.975}_{(8,\,7)} = 4.90$$

$$F^{\alpha/2}_{(n_1-1,\, n_2-1)} = F^{.025}_{(8,\,7)} = \frac{1}{F^{.975}_{(7,\,8)}} = \frac{1}{4.53} = 0.22$$

وبناء على الحالة الاولى التي بالجدول 2-5 يكون القرار برفض H_0 اذا كانت $F^* > 4.90$ أو $F^* < 0.22$ وبما أن $F^* = 0.44$ موجودة بين القيمتين الجدوليتين أي تقع في منطقة القبول، فبالتالي لا نرفض H_0 وهذا ما يدعم صحة الافتراض الاساسي لاختبار t المستخدم في المثال 2-1.

تحليل التباين

ANALYSIS OF VARIANCE

3 - 1 مقدمة :

درسنا في الفصل السابق اختبار الفرضيات التي تقارن بين متوسطي مجتمعين . ولكن في الواقع هناك تجارب كثيرة ومسائل إحصائية تتعلق باختبار متوسطات ثلاثة مجتمعات فاكثر . فمثلاً إذا كان لدينا أربعة أنواع من المبيدات ونريد معرفة ما إذا كانت المبيدات الأربعة تعطي نفس النتائج أم أن هناك إختلافاً معنوياً بينها . فبإمكاننا استخدام اختبار t لمقارنة متوسطي مجتمعين لكل زوج من المجتمعات الأربعة . وهذا يعني القيام بستة اختبارات t . فإذا كانت متوسطات المجتمعات الأربعة كالآتي : $\mu_1, \mu_2, \mu_3, \mu_4$ تكون الفرضيات المطلوب اختبارها هي :

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

والعيب الرئيسي لهذه الطريقة هو أنها ليست عملية ، حيث يزداد عدد المقارنات بسرعة كلما ازداد عدد المجتمعات ، فبالنسبة لعدد k من المتوسطات يكون عدد المقارنات الزوجية هو $C_2^k = \frac{k!}{2!\,(k-2)!}$. فمثلاً إذا كان عدد المتوسطات k = 6 فهذا يعني أننا نحتاج للقيام باختبار t 15 مرة. أما العيب الثاني البارز لهذه

الطريقة هو أن احتمال رفض فرضية H_0 واحدة على الأقل وهي صحيحة ، أي رفضها خطأ، يزداد كلما ازداد عدد اختبارات t . وبالتالي وإن حدد مستوى المعنوية ، أو احتمال رفض H_0 وهي صحيحة ، عند مثلاً $\alpha = .05$ لكل اختبار t فإن احتمال رفض فرضية واحدة على الأقل خطأ يزيد عن 0.05 . وبعبارة أخرى فإن الاحتمال المشترك للوقوع في خطأ من النوع الأول للاختبارات الستة يكون أكبر من $\alpha = .05$. فبالنسبة للاختبار الواحد يكون احتمال الوصول إلى القرار الصحيح ، أي قبول H_0 وهي صحيحة، هو 0.95 ، وبالتالي فاحتمال الوصول إلى القرار الصحيح للاختبارات الستة هو $(0.95)^6$. ويعني هذا أن احتمال الوصول إلي قرار واحد على الأقل غير صحيح هو $1 - (0.95)^6 = 0.265$.

وفي الواقع وبالنسبة لهذه التجربة المطلوب هو اختبار فرضية واحدة وهي :

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

أي هل المتوسطات الأربعة متساوية أم لا ؟ وعكس فرض العدم هو الفرض البديل التالي :

H_a : متوسط واحد μ_i على الأقل يختلف عن الباقي .

والاختبار المناسب لفرض العدم هو اختبار F الذي نحصل عليه باستخدام طريقة تحليل التباين . ويعتبر تحليل التباين من أقوى الطرق الإحصائية المتوفرة في علم الإحصاء وأكثرها شيوعاً في البحوث التجريبية .

تتلخص طريقة تحليل التباين في عملية حسابية تجري على البيانات وذلك لتجزئة مجموع مربعات الانحرافات الكلية للاستجابة Y إلى عدد من المجاميع المختلفة طبقاً للمصادر المسببة للاختلاف . وتقسم درجات الحرية الكلية أيضاً طبقاً للمصادر نفسها . ومن هنا يفي تحليل التباين بالهدف الأول للتجارب العلمية والذي هو الإجابة عن السؤال: هل هناك فروق بين المتوسطات ؟ (الفصل الأول) . وأبسط تحليل للتباين هو الذي يطبق على البيانات المصنفة حسب

ظاهرة واحدة أو بيانات أحادية التقسيم . وسنتطرق لهذا التحليل في الفقرة التالية .

3 - 2 تحليل التباين الأحادي One way ANOVA

في معظم ميادين البحث العلمي يقوم الباحث بإجراء تجربة على عدد من الوحدات التجريبية فيعاملها بمعالجات مختلفة لدراسة آثارها على المادة التجريبية . وفي آخر التجربة يقوم بجمع المشاهدات وتصنيفها إلى مجموعات حيث توضع كل مجموعة تحت المعالجة الخاصة بها. ولهذا تكون البيانات قد صنفت على أساس خاصية واحدة وهي المعالجات بحيث تمثل كل مجموعة عينة عشوائية مسحوبة من مجتمع المعالجة . والمراد هنا هو اختبار ما إذا كانت متوسطات المجتمعات متساوية أم لا ؟ وذلك بواسطة تلك العينات المسحوبة . وهذا هو الأمر السائد في الإحصاء الاستدلالي وهو سحب عينة للاستدلال بها عن مجتمع أو عينات للاستدلال بها عن مجتمعات عدة .

ولتوضيح العمليات الحسابية في تحليل التباين يمكن ترتيب البيانات كما في الجدول (1-3) . ونعبر بالرمز Y_{ij} عن المشاهدة j من المجموعة i بحيث $i = 1,...,k$ و $j = 1,...,r$. وفيما يلي نعرف بعض الرموز المعتمدة في الإحصاء للتعبير عن المجاميع والمتوسطات :

المجموع والمتوسط العام	المتوسطات	المجاميع
	$\overline{Y}_{1.} = Y_{1.}/r$	$Y_{1.} = \sum_{j=1}^{r} Y_{1j}$
$Y_{..} = \sum_{i=1}^{k}\sum_{j=1}^{r} y_{ij}$	$\overline{Y}_{i.} = Y_{i.}/r$	$Y_{i.} = \sum_{j=1}^{r} Y_{ij}$
:	:	
$\overline{Y}_{..} = Y_{..}/kr$	$\overline{Y}_{k.} = Y_{k.}/r$	$Y_{k.} = \sum_{j=1}^{r} Y_{kj}$

جدول (3-1) : بيانات التصنيف الأحادي

المشاهدات	المجموعات (i) 1	2	...	k	
1	Y_{11}	Y_{21}	...	Y_{k1}	
(j) 2	Y_{12}	Y_{22}	...	Y_{k2}	
:	:	:	...	:	
r	Y_{1r}	Y_{2r}	...	Y_{kr}	
المجموع	$Y_{1.}$	$Y_{2.}$	...	$Y_{k.}$	$Y_{..}$
المتوسط	$\overline{Y}_{1.}$	$\overline{Y}_{2.}$	...	$\overline{Y}_{k.}$	$\overline{Y}_{..}$

الافتراضات :

لسلامة اختبار F في تحليل التباين يكون من الضروري توفر افتراضين أساسيين على غرار ما تم افتراضه في اختبار t لمقارنة متوسطي مجتمعين، ولكن في تحليل التباين ستمد تلك الافتراضات إلى k متوسط بحيث تكون كالتالي :

1 - أن $Y_{ij} \sim N(\mu_i, \sigma_i^2)$ مستقلة عن بعضها البعض $i=1,\dots, k\ ;\ j = 1,\dots, r$

2 - أن $\sigma^2 = \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$

وتكون الفرضيات كما يلي :

فرض العدم : $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$

الفرض البديل : H_a : μ_i واحد على الأقل يختلف عن الباقي

جدول تحليل التباين :

لاختبار فرض العدم نقوم بتجزئة مجموع المربعات الكلية إلى قسمين : أحدهما هو مجموع مربعات داخل المجموعات والآخر مجموع مربعات بين المجموعات . ونستخدم مقارنة هذين الجزءين ، بعد قسمتها على درجات الحرية الخاصة بها ، للحكم على فرض

العدم بالقبول أو بالرفض . ومجموع المربعات هو :

- مجموع المربعات الكلية :

$$SSTo = \sum_{i=1}^{k}\sum_{j=1}^{r}(Y_{ij} - \overline{Y}_{..})^2$$

$$= r\sum_{i=1}^{k}(\overline{Y}_{i.} - \overline{Y}_{..})^2 + \sum_{i=1}^{k}\sum_{j=1}^{r}(Y_{ij} - \overline{Y}_{i.})^2$$

$$= \quad SSB \quad + \quad SSW$$

حيث : SSB هو مجموع المربعات بين المجموعات و SSW هو مجموع المربعات داخل المجموعات

وتجزأ درجات الحرية الكلية كما يلي :

$$df_{SSTo} = df_{SSB} + df_{SSW}$$

$$kr - 1 = k - 1 + k(r - 1).$$

وبإمكاننا حساب مجموع المربعات بطريقة حسابية مبسطة ، وذلك لاستخدام الآلة الحاسبة ، على النحو التالي :

$$SSTo = \sum_{i=1}^{k}\sum_{j=1}^{r} y_{ij}^{2} - CF \ , \ CF = (Y_{..})^2/k\,r \qquad \text{(Correction factor)}$$

$$SSB = \frac{1}{r}\sum_{i=1}^{k} Y_{i.}^{2} - CF$$

$$SSW = SSTo - SSB$$

ومن مجموع المربعات تحسب متوسطات المربعات كالتالي :

$$MSB = SSB / (k - 1)$$

$$MSW = SSW / k(r - 1)$$

ثم ننهي حسابات جدول تحليل التباين بإيجاد قيمة F المحسوبة

$$F = \frac{MSB}{MSW}$$

والتي على اثرها يتم الحكم على فرض العدم H_0 بالرفض عند مستوى المعنوية α إذا كان

$$F > F^{1-\alpha}_{k-1,\, k(r-1)}$$

ونلخص جدول تحليل التباين في الجدول (3-2)

جدول (3-2) : جدول تحليل التباين الأحادي .

مصدر الاختلاف Source of variation	درجات الحرية Degrees of freedom	مجموع المربعات Sum of squares	متوسطات المربعات Mean squares	قيمة F
بين المجموعات Between groups	k -1	SSB	MSB	$F = \frac{MSB}{MSW}$
داخل المجموعات Within groups	k(r-1)	SSW	MSW	
المجموع Total	kr - 1	SSTo		

مثال (3-1) :

تمثل البيانات التالية محصول (كلغ) 5 أصناف مختلفة من البرتقال (A,B,C,D,E) ، ومن كل صنف سحبت عينة عشوائية متكونة من سبع أشجار وسجلت البيانات في الجدول (3-3) .

وتحسب مجموع المربعات على النحو التالي :

$$SSTo = \Sigma\Sigma Y_{ij}^2 - CF \quad , \quad CF = \frac{Y_{..}^2}{k\,r} = \frac{(552)^2}{5(7)} = 8705.83$$

$$= 9720 - 8705.83 = 1014.17$$

$$SSB = \frac{1}{r}\sum_{i=1}^{5} Y_{i.}^2 - CF$$

$$= \frac{1}{7}[84^2 + 118^2 + ... + 112^2] - CF = 9059.14 - 8705.83 = 353.31$$

جدول (3-3) : محصول (كلغم) خمسة أصناف من البرتقال .

أصناف	A	B	C	D	E	
المشاهدات	7	13	20	9	18	
	9	15	22	14	16	
	20	16	21	21	17	
	19	14	27	22	14	
	11	22	23	8	22	
	6	22	15	10	10	
	12	16	19	7	15	
$Y_{i.}$ المجموع	84	118	147	91	112	$Y_{..} = 552$
$\overline{Y}_{i.}$ المتوسط	12.0	16.86	21.0	13.0	16.0	$\overline{Y}_{..} = 15.77$
S_i^2 التباين	30.67	13.48	13.67	38.67	13.67	$\Sigma\Sigma Y_{ij}^2 = 97.20$

$$SSW = SSTo - SSB = 1014.17 - 353.31 = 660.86$$

$$MSB = SSB / (k - 1) = 353.31/4 = 88.328$$

$$MSW = SSW / k(r-1) = 660.86 / 5(7-1) = 22.029$$

$$F = MSB / MSW = 88.328 / 22.029 = 4.01$$

وتلخص كل هذه الحسابات في جدول تحليل التباين الأحادي الموضح في الجدول (3-4). أما F الجدولية عند مستوى المعنوية $\alpha = .05$) فهي $F^{1-\alpha}_{k-1,\, k(r-1)} = F^{.95}_{4,30} = 2.69$ وبما أن قيمة F المحسوبة ، $F = 4.01$ ، هي أكبر من F الجدولية فنرفض H_0 وتعتبر أن هناك فروقاً معنوية بين محاصيل أصناف البرتقال الخمسة . وطبعاً وصلنا إلى هذا القرار بناء على ما توفر لدينا من بيانات ، لذلك فمن الأفضل صياغة القرار على النحو التالي : بناء على هذه البيانات هناك دليل قوي على عدم صحة فرض العدم H_0 .