وزارة التعليم العالي و البحث العلمي
جامعة طاهري محمد بشار
كلية العلوم اقتصادية التجارية و علوم التسيير
قسم العلوم مالية ومحاسبة

النماذج والأساليب الإحصائية لتحليل البيانات

حسب المقرر الرسمي لوزارة التعليم العالي و البحث العلمي

من اعداد الدكتورة عيشوش خيرة

السنة الجامعية 2020-2021

فهرس

مقدمة :

كان ولازال علم الإحصاء و تحليل البيانات أحد الركائز الأساسية في اتخاذ القرارات بمختلف أنواعها وعلى مختلف المستويات الفردية الشخصية و المؤسساتية، وواحدة من الأدوات المحورية في المجال البحثي الأكاديمي التي تبنى على اساسها النتيجة و القرار و مع تزايد الاهتمام في مجال علوم البيانات في عصرنا الحالي أصبح من المهم لكل فرد أكاديمي أو مهني الإلمام بالمفاهيم الاساسية في علم الإحصاء لتوظيفها و استخدامها بشكل صحيح في مختلف القطاعات و الاختصاصات.

وتحليل البيانات هو عبارة عن استخدام للوسائل الحسابية و الرياضية في تجميع البيانات ومن ثم تنظيمها و تبويبها بغرض وصفها و تفسيرها للظاهرة و فهم مختلف العلاقات فيما بينها بشكل يساعد في الوصول إلى تحقيق هدف الدراسة.

وانطلاقا من هذه الأهمية التي يكتسيها علم تحليل البيانات و بناء النماذج الإحصائية جاءت هذه المطبوعة بهدف تقديم و توضيح وشرح المفاهيم الأساسية في علم الإحصاء وتحليل البيانات و,lr$ابناء نماذج إحصائية ، و هي موجهة إلى طلبة العلوم اقتصادية و التجارية و علوم التسيير من أجل تمكينهم من تحليل المعطيات و بناء النماذج الكمية وفهم أكثر للمقياس و قد تم تكيفها حسب المقرر الرسمي لوزارة التعليم العالي و البحث العلمي.

تتألف هذه السلسة من ستة فصول متكاملة و متتابعة:

- **الفصل الأول**:مقدمة في تحليل البيانات الإحصائية.
- **الفصل الثاني** : تحليل التباين.
- **الفصل الثالث** : تحليل الارتباط.
- **الفصل الرابع** : تحليل الانحدار .
- **الفصل الخامس** : تحليل السلاسل الزمنية.
- **الفصل السادس** : التحليل العاملي.

1- مفهوم علم الإحصاء :

يعرف علم الإحصاء بأنه العلم الذي يختص بالطرق العلمية لجمع البيانات وتنظيمها وتلخيصها وعرضها بيانياً ثم تحليلها وتفسيرها وإجراء المقارنات واستنتاج العلاقات بهدف استخدامها في اتخاذ القرارات المناسبة.

ويعرف علم الإحصاء بأنه فرع من فروع الرياضيات يشمل النظريات والطرق الموجهة نحو جمع البيانات، وصف البيانات، الاستقراء، صنع القرارات، ويتميز باستخدام الأرقام والرموز والدوال الرياضية، والمقاييس والجداول والرسوم البيانية.

2- تقسيمات علم الإحصاء :

لقد بدأ الإحصاء كعلم وصفي بحث ولكنه تطور الى أداة قوية لاتخاذ القرار مع نمو فرع الاستدلال فيه وأصبح علم الإحصاء ينطوي على فرعين رئيسيين يكمل كل منهما الآخر هما :

- **الإحصاء الوصفي :**

يتضمن عملية جمع البيانات وترتيبها في جداول وتمثيلها في رسوم بيانية، ومنحنيات، وأشكال توظيفية تساعد على وصف الظواهر الانسانية والاجتماعية، كما يتضمن الكشف عن مدى تجمع البيانات العددية و تشتتها والارتباط بينهما.

والإحصاء الوصفي هو الإحصاء الذي يشتمل على مجموعة من المبادئ الإحصائية التي تساعد على فهم الظاهرة المدروسة مثل المتوسطات والانحرافات المعيارية.

- **الإحصاء الاستدلالي أو التحليلي أو الاستنباطي :**

ويشتمل على الطرق الإحصائية التي تستخدم للوصول الى قرارات وأحكام واستنتاجات عن المجتمع المدروس باستخدام عينة مسحوبة منه وذلك من خلال دراسة الفروق بين المتوسطات **(T.Test)** تحليل التباين **(Anova)** لمقارنة الفروق بين أكثر من متوسطيين، معاملات الارتباط بيت متغيرات الدراسة...الخ.

3- المصطلحات الضرورية في التحليل الإحصائي:

1-3 مجتمع وعينة الدراسة :

- ✓ **المجتمع :** مجتمع البحث أو الدراسة هو المجموعة الكلية من المفردات أو العناصر التي يهتم بها الباحث وتعمم نتائجه عليها، والمفردات أو العناصر قد تكون أشخاصاً أو اسم أو

مؤسسات أو مراكز أو مصانع أو غيرها، فمثلاً طلبة جامعة أو موظفو شركة، أو سكان مدينة كلها أمثلة ممكنة عن المجتمع.

- ✓ **العينة** : أما العينة فهي جزء صغير من المجتمع أو المجموعة الكلية(المجتمع) تجري عليها التجربة أو التطبيق، إذا تم التطبيق على العينة وتعمم النتائج على المجتمع، والسبب في اللجوء الى العينة بدلاً من التطبيق على مجتمع الدراسة كاملاً أن مثل هذا التطبيق فيالغالب يتطلب المثير من الوقت والجهد والمال، بالإضافة الى صعوبة التطبيق على جميع أفراد المجتمع في الكثير من الأحيان.

أما خطوات اختيار العينة فهي كما يلي:

- ✓ **تحديد المجتمع المقصود** :وهو كما ذكرنا المجتمع الذي يهتم به الباحث والذي تعمم عليه نتائج البحث.
- ✓ **تحديد وحدة العينة** : ووحدة العينة مفردة غير قابلة للتقسم.
- ✓ **تحديد اطار العينة** : واطار العينة هو الافضل أو أكبر قائمة متوفرة بأسماء أفراد المجتمع.
- ✓ **تحديد حجم العينة** : يحدد مدى تجانس المجتمع حجم العينة، فكلما كان التجانس أكبر قل حجم العينة المطلوب، فمثلاً حجم العينة المطلوب من مجتمع متفاوت فيه مثلاً ذكور وإناث صغار و كبار وعرب وأجانب يكون أكبر من حجم العينة المطلوب من مجتمع أكثر تجانساً فيه طلاب ذكور فقط ومن نفس العمر والثقافة، و على العموم كلما زاد حجم العينة كلما زادت فرصة تمثيل المجتمع بصورة أدق وقل خطأ القياس، ومن العوامل المهمة في تحديد الحجم الأدنى المطلوب للعينة نوع الاختبار الإحصائي المستخدم في التحليل فمثلاً تحتاج الى **30** فرداً على الأكثر في كل مجموعة في اختبار « **T** »المستقل، و الى **30** فرد على الأقل في اختبار تحليل التمايل و نحتاج الى **10** أفراد على الاقل مقابل كل فقرة من فقرات الاختبار أو الاستبيان المستخدم في التحليل العاملي.

4- أنواع العينات حسب طريقة اختيارها :

العينات الاحتمالية : وهي العينات التي يخضع اختيارها لقوانين احتمالية وأهمها ما يلي : (الحصادي ونصار وعتيلي، **2005**).

1. **العينة العشوائية البسيطة (Simple Randon Sample)**: و فيها يكون لكل فرد في المجتمع نفس فرصة الاختيار، واختيار فرد ما لا يؤثر على فرصة اختيار فرد أخر، ويتم اختيار العينة

من خلال إعداد قائمة بأفراد أو عناصر المجتمع، وإعطاء رقم لكل فرد في القائمة، ثم يحدد حجم العينة المطلوب، و بعدها يتم اختيار أفراد العينة عشوائياً (بواسطة الكمبيوتر مثلا)، وهذا هو الأسلوب الشائع في اختيار الفائزين في المسابقات والفائزين في سحوبات المراكز التجارية وهو الاسلوب الأمثل إن أمكن استخدامه.

2. **العينة المنتظمة** (SystematicSample): ينصح باستخدام هذا الأسلوب عندما لا تتوفر قائمة بأفراد مجتمع الدراسة، مثال ذلك، إذا أراد أحد الباحثين أن يدرس جودة الطعام في أحد المطاعم فإنه من الصعب عليه أن يحصل على قائمة بأسماء جميع زبائن المطعم، ولكنه يمكن أن يقف على باب المطعم ويختار شخصاً ما عشوائياً من الداخلين إليه ليبدأ به وبعدها يختار الشخص التالي بعد دورة محددة الطول، فمثلاً لو إختار عشوائيا الشخص السابع وحدد طول الدورة عشوائياً و ليكن 10 مثلاً، فإن الاشخاص الذين سيختارهم بهذه الطريقة هم : السابع عشر، والسابع و العشرين، والسابع والثلاثين، والسابع والأربعين، حتى يتجمع لديه العدد الكافي حسب حجم العينة المطلوب المحدد مسبقاً.

3. **العينة الطبقية** (Stratified Sample) :قد لا يوفر الاختيار العشوائي عينة تمثل الخصائص الفعلية للمجتمع وخصوصاً إذا كان فيه مجموعات أو طرقات متباينة من حيث تأثيرها في تجربة البحث، عندها يفضل أن يتم الاختيار عشوائياً ولكن من كل طبقة أو مجموعة في المجتمع على حده، ويفضل اختيار عدد من الأفراد في كل مجموعة حسب تواجد تلك المجموعة في المجتمع الأصلي فمثلاً إذا كانت نسبة الإناث في مجتمع ما **60%** و نسبة الذكور **40%**، ولم تعكس العينة العشوائية هذه النسب بشكل دقيق، فيمكن عندها تخصيص عينة للذكور وأخرى للإناث بحيث تعكس نسب تواجد الجنسين في المجتمع بشكل دقيق.

4. **العينة العنقودية** (Cluster Sample) : يستعمل هذا الاسلوب عندما تتوفر تجمعات طبيعية واضحة متجاورة مكاناً أو زماناً ضمن مجتمع واحد، كالصفوف الدراسية في الجامعة، وتسمى هذه التجمعات عناقيد، فمثلاً بدلا من اختيار **600** طالب جامعي بشكل عشوائي فردي، يتم اختيار عدد من الصفوف الدراسية في وقت محدد عشوائياً بحيث يكون العدد الكلي للطلبة مسارياً للعدد المطلوب، فإذا افترضنا أن معدل عدد الطلبة في الصفوف الدراسية في تلك الجامعة **30** طالباً، فإننا بحاجة الى **20** صفاً حتى يكون لدينا **600** طالباً، فيتم اختيار

الصفوف ال **20** عشوائياً من القائمة الكلية للصفوف الدراسية في الجامعة، ويشترك في الدراسة أو البحث جميع طلبة الصف المختار بهذه الطريقة.

- **العينات غير الاحتمالية** : لا تخضع طريقة الاختيار هنا لقوانين الاحتمالات، وبالتالي لا تعمم نتائج البحث على المجتمع الذي اختيرت منه العينة بسبب التحيز في اختيارها وعدم تمثيلها للمجتمع، ولكن اختيار عينة غير احتمالية قد يكون لأسباب عملية منها صعوبة الحصول على عدد كاف أو صعوبة تحديد مجتمع الدراسة وإطار العينة كما هو الحال في البحوث ذات الطبيعة الحساسة كتعاطي المخدرات مثلا، كما قد لا تتوفر للباحث الإمكانات المادية لاختيار عينة ممثله، فيلجأ عندئذ الى هذه الطرق غير الاحتمالية، ومنها تطبيق البحث على متطوعين أو على ما يتوفر لدى الباحث من أفراد كما هو الحال في العينة الميسرة أو المتوفرة، أو أن يختار الباحث عينة محددة بناء على معطيات و مبررات عملية كما هو الحال في العينة المقصودة.

5- ثبات أداة القياس :

مفهوم ثبات أداة القياس.

قبل إجراء البحوث واختيار الفرضيات فإنه لا بد من التأكد من موثوقية أداة القياس المستخدمة. حيث تعكس الموثوقية هنا درجة ثبات أداة القياس.

وتتأثر درجة ثبات أداة القياس بعدة عوامل أهمها:

1. **طول الاختيار** : تزداد قيمة معامل الثبات بزيادة عدد العبارات في الاستبيان حيث يؤثر زيادة عدد العبارات على شمولية المحتوى.
2. **تجانس أو تباين درجات أفراد العينة** :يقل معامل ثبات الاختيار بزيادة تجانس أفراد العينة. ويكبر معامل الثبات مع زيادة تباين أفراد العينة في اجاباتهم.
3. **مدى صعوبة فهم أداة القياس** : عندما تكون عبارات الاستبيان غير مفهومة أو صعبة الاستيعاب فإن أفراد العينة المبحوثين قد يلجأوا الى التخمين وبالتالي تتأثر درجة ثبات أداة القياس.

4. **الفترة الزمنية بين قياسين بنفس الأداة** : إذا كانت الفترة الزمنية بين قياسين بنفس الأداة طويلة نسبياً فإن الظروف قد تختلف وبالتالي قد يؤثر ذلك على اجابات بعض أفراد العينة المبحوثين مما يؤدي الى عدم ثبات القياس.

هناك نوعين من الثبات في مجال أداة القياس :

1. **الثبات الداخلي** : المقصود بالثبات الداخلي مدى اتصاف عبارات القياس بالتناسق الداخلي. وهناك عدة مقاييس لاختيار الثبات الداخلي للأداة من أهمها معامل كروباخ ألفا **Cronbach's alpha** و التجزئة النصفية **Split half**.
2. **الثبات الخارجي** : والذي يتعلق بدرجة ثبات أداة القياس بمرور الوقت ويمكن قياس الثبات الخارجي من خلال تطبيق نفس أداة القياس مرتين وعلى فترتين متقاربتين ثم قياس معامل الارتباط بين نتائج المرة الأولى ونتائج المرة الثانية.

6- تعريف المتغير و أنواع المتغيرات :

يعرف المتغير على أنه أي شيء يكون له عدد من القيم المختلفة في الوقت نفسه مثل تحصيل الطلبة. وأرباح الشركات، أو يكون متغيراً عبر الزمن مثل إدراك الفرد لظاهرة ما طبقا لمراحل نموه. ويمكن أن يحتل المتغير أي موضع على مقياس متصل ويسمى المتغير في هذه الحالة متغيراً متصلاً، ولكن قد يأخذ المتغير قيم محددة مثل متغير نوعي ويسمى في هذه الحالة متغيراً منفصلاً.

وتقسم المتغيرات بشكل أساسي الى :

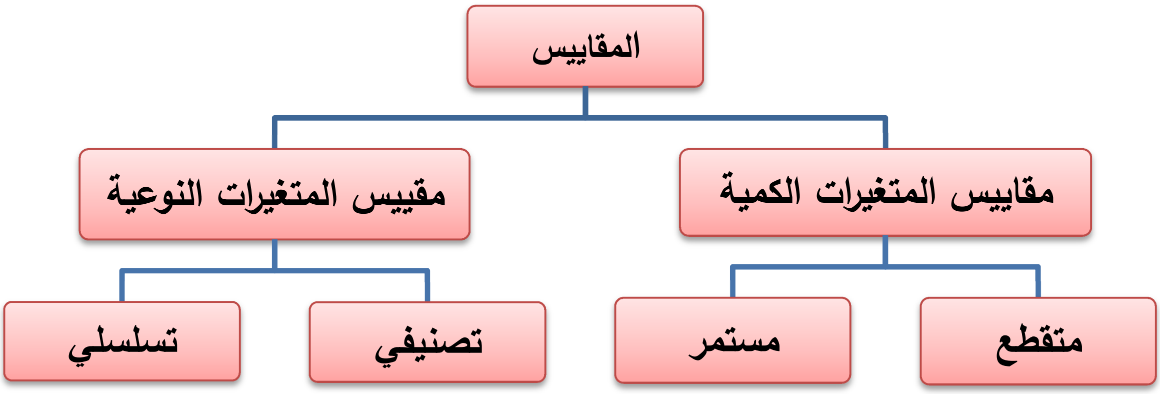
- **أولا: متغير مستقل Independent Variable** : وهو المتغير الذي يؤثر على المتغير التابع ولا يتأثر به.
- **ثانيا : متغير تابع Dépendent Variable** :وهو المتغير الذي يتأثر بتغير المتغير المستقل.

مثال : دراسة أثر الإعلان على مبيعات الشركات، فالإعلان في هذه الحالة هو متغير مستقل والمبيعات متغير تابع. وكذلك الحال بالنسبة لأثر جنس الطالب على التحصيل العلمي، فالجنس متغير مستقل و التحصيل العلمي متغير تابع.

ويمكن أن يظهر المتغير مستقلا في دراسة ما وتابعا في دراسة أخرى، فمثلا إذا درسنا أثر الحوافز المادية على الرضا الوظيفي فالرضا الوظيفي يكون متغيراً تابعا والحوافز مستقلة. لكن إذا درسنا أثر الرضا الوظيفي على الولاء التنظيمي فيكون الرضا في هذه الدراسة متغيراً مستقلاً.

7- أنواع البيانات والمقاييس الاحصائية :

إن قياس المتغيرات يعتمد بشكا أساسي على نوع المتغير والذي يمكن تقسيمه طبقا للشكل **(1-1)** الى الأنواع التالية :



(1−1) أنواع المقاييس الإحصائية

البيانات النوعية :

نحصل على هذا النوع من البيانات عندما تكون السمة تحت الدراسة هي سمة نوعية، والتي يمكن تصنيفها بحسب أصناف أو أنواع وليس بمقاييس عددية، مثلا يصنف الجنس الى (الذكر وأنثى). ويندرج تحت هذا النوع من البيانات المقياس التصنيفي، والمقياس التسلسلي.

1. **المقياس التصنيفيNominal Scale** : وهذا المقياس يمكن الباحث من تجزئة و تصنيف المتغيرات الى مجموعات كيفيات. مثال ذلك متغير الجنس ويتم تصنيفه الى مجموعات هي (ذكر، وأنثى) ويتم إعطاء رمز **(1)** للدلالة على مجموعة الذكور، ورمز **(2)** للدلالة على مجموعة الإناث، وهذا العدد ليس له قيمة بل رمز للتفريق بين مجموعات كيفيات، و المقياس التصنيفي يساهم في بيان تكرار ونسبة كل فئات $[B.A[$, $].A[$, $]B.A]$, $[B.A]$ تكون للمتغير مستمر كمي.
2. **المقياس التسلسلي Ordinal:** وهذا المقياس يتعدى تصنيف المتغيرات بشكل يدل على الفروق النوعية بين المجموعات الى ترتيب كيفيات بشكل ذي معنى أو مغزى، فهو يعمل على ترتيب كيفيات ضمن المتغيرات التي يمكن ترتيبها بحسب الأفضلية/ أي من الأفضل الى الأسوأ بإعطائها رقم **3،2،1....**

مثال :على افتراض أن باحثاً يرغب بتحديد العوامل التالية التي يفضلها المستهلك عند شراء سلعة ما وذلك بإعطاء رقم (1)للمهم ورقم(2)للتالي وهكذا :

( ) السعر.

( ) خدمات ما بعد البيع.

( ) الجودة.

( ) مكان الصنع.

والمقياس التسلسلي يساعد على تحديد النسبة المئوية للمجيبين الذين يعتبرون السعر الأكثر أهمية، وخدمات ما بعد البيع هي الأكثر أهمية وهكذا. من هنا نلاحظ أن المقياس التسلسلي يبين الاختلاف بين المجموعات ويعطي معلومات حمل كيفية تمييز المجيبين بين العناصر والمجموعات المختلفة من خلال ترتيبها.

إلا أن المقياس التسلسلي لا يبين حجم الفروق بين الرتب المختلفة، فمثلا إذا كان الفرق بين الترتيب الأول والثاني قليلاً والثاني و الثالث كبيراً فالمقياس التسلسلي لا يظهر حجم هذه الفروق لأنه يقوم على إعطاء أرقام متسلسلة دون بيان حجم الفروق.

البيانات الكمية :

نحصل على هذا النوع من البيانات عندما تكون السمة السائدة تحت الدراسة قابلة للقياس على مقياس عددي، لذا فالبيانات التي نحصل عليها تتألف من مجموعة أعداد، وتسمى بيانات كمية أو عددية ويندرج تحت هذا النوع من البيانات المقياس الفئوي، والمقياس النسبي.

1. **المقياس الفئوي IntervalScale**: وهذا المقياس يسمح للباحث بإجراء العمليات الحسابية على البيانات المجمعة من قبل المبحوثين، فهو يسمح باستخدام مقاييس النزعة المركزية ومقاييس التشتت. و المقياس الفئوي يفرق الأفراد على تصنيفات معينة كما هو الحال في المقاييس السابقة إلا أنه يبين حجم الفروق بين الأفراد، و يفتقر هذا المقياس الى ن قطة البداية صفر مثل درجة الحرارة فهي عندما تكون درجة الحرارة صفر فنتوقع أن تكون لا حرارة نهائيا – وهذا غير صحيح– بالتالي فنقطة البداية هي نقطة عشوائية أو افتراضية، ومن المقاييس الفئوية شائعة الاستخدام :

- **أولا : مقياس ليكرت.**

ويعتبر مقياس ليكرت أكثر المقاييس شيوعاً لقياس الاتجاهات والذي يمكن أن يكون مقياس ثلاثي أو خماسي أو سباعي كما هو مبين في الشكل التالي :

أ- مقياس ليكرت ثلاثي :

المقياس	أوافق بشدة	محايد	لا أوافق بشدة
العلامة	3	2	1

2. مقياس ليكرت الخماسي :

المقياس	أوافق بشدة	أوافق	محايد	لا أوافق	لا أوافق بشدة
العلامة	5	4	3	2	1

3. مقياس ليكرت سباعي :

المقياس	أوافق بشدة					لا أوافق بشدة
العلامة	7	6	5	4	3	2	1

- **ثانيا : مقياس التمثيل البياني.**

ويستخدم لتوضيح إجابات عن سؤال معين عن طريق وضع علامة على نقطة ضمن النقاط الموجودة على الخط البياني، كما هو مبين في الشكل التالي :

➢ **ثالثا : مقياس فروق المعاني.**

ويشبه هذا المقياس مقياس ليكرت السباعي لكن بدلا من استخدام أوافق بشدة، يتم إعطاء كلمات مرادفة و يختار المستجيب درجة بين الكلمتين المترادفتين. ويتم إعطاء المقياس الأوزان **3–** الى **3+** أو **1** الى **7** مثلا :

ممتاز ------------------- سيء

7 6 5 4 3 2 1

أو

3 2 1 0 –1 –2 –3

واسع --------------------- ضيق

7 6 5 4 3 2 1

مناسب ------------------ غير مناسب

7 6 5 4 3 2 1

➢ **رابعا : مقياس التمثيل من خلال التعداد.**

وفي هذا المقياس يتم استخدام عدد من البدائل، والتي من خلالها يتم اختيار البديل الأكثر تمثيلا للموضع الذي يناسب الفرد مثل :

العبارة	مهم جدا	مهم	مهم الى حد ما	غير مهم إطلاقا
يشكــل السعر عاملا مهما عند اقتناء المنتج	4	3	2	1

3. المقياس النسبي Radio Scale: يتغلب هذا المقياس على شكل نقطة البداية فنقطة البداية في هذا المقياس تكون صفراً. لذا فهو يحسب الفرق بين نقطتين، ويبين الاختلاف بين هاتين النقطتين، بالتالي فهو يسمح بتحديد نسبة الاختلاف بين الأفراد ممثلاً يعتبر عمر شخص

50 سنة ضعف عمر شخص عمره 25 سنة، وكذلك يسمح بإجراء العمليات الحسابية مثل مقاييس النزعة المركزية ومقاييس التشتت ومن الامثلة على هذا المقياس العمر، الدخل.

8. مصادر جمع البيانات :

تقسم مصادر جمع البيانات الى مصدرين هما المصدر الأولية والمصدر الثانوية :

أ- **المصدر الأوليةPrimarySource:** وهذه المصادر تشكل مادة الأولية التي يتم جمعها من خلال عملية البحث. وهذه البيانات يتم جمعها من خلال الباحث نفسه باستخدام عدد من الطرق منها الاستبانة، و المقابلة والملاحظة. وتتصف هذه المصادر بالموثوقية لكون الباحث على اطلاع بكيفية جمع البيانات ومصادرها.وتشمل المصادر الأولية:

1. الملاحظةObservation:

تعتبر الملاحظة من أدوات جمع البيانات حيث تستخدم عادة في مجال دراسات الطبيعة والسلوك الانساني. وهناك عدة تصنيفات للملاحظة من أهمها التصنيف القائم على أساس دور الباحث والذي يقسم الملاحظة الى نوعين أساسيين :

أ- **الملاحظة المشاركة :** حيث يشارك الباحث المبحوثين حياتهم ومشاكلهم ومناقشاتهم ويعيش معهم لحظة بلحظة.

ب- **الملاحظة غير المشاركة :** وهي أكثر أنواع الملاحظات انتشاراً حيث يجلس الباحث في مكان معين ليلاحظ ويراقب سلوكالمبحوثين بدون أن يشعروا بأنه يتولى مراقبتهم.

وبالرغم من وجود مزايا عديدة لاستخدام الملاحظة كأداة لجمع البيانات من أرض الواقع وعدم تأثر البيانات بمزاجية المبحوثين أوشخصياتهم، إلا أن هناك بعض العيوب التي تصاحب عملية الملاحظة مثل أخطاء تفسير السلوك الملاحظ و الفروقات الفردية في ادراك الأمور و تسجيلها لدى الذين يقومون بعملية الملاحظة.

2. المقابلة Interview :

المقابلة هي تفاعل لفظي بين شخصين أو أكثر من خلال حوار كلامي وجها لوجه أو من خلال وسائل أخرى مثل الهاتف أو الأقمار الصناعية. ويقوم الباحث في المقابلة بدور المقابل أي الذي يجري المقابلة حيث يوجه بعض الأسئلة والاستفسارات الى الطرف الآخر الذي تجري معه المقابلة والذي يقوم بدوره بإجابة الأسئلة و الرد على الاستفسارات المقدمة.

وعند قيام جامع البيانات بإجراء المقابلة فلابد له من توفير الجو الودي منذ بدء عملية المقابلة وحتى انتهاؤها. كذلك لا يجوز لجامع البيانات محاولة فرض وجهة نظره أو رأيه في أي موضوع مع مراعاة بساطة الأسئلة الموجهة وسهولة فهمها، كما ينبغي التأكيد على تنظيم وقت المقابلة بحيث تنتهي الأسئلة مع اجاباتها في الوقت المحدد.

تتمتع المقابلة بمزايا عديدة من حيث أنها توفر امكانية توضيح بعض الأمور واعادة طرح الأسئلة والاستفسارات مرة أخرى. بالإضافة الى أنها تتضمن مؤشرات معينة تعزز الإجابات مثل نغمة الصوت وملامح الوجه عند الإجابة وحركة اليدين، إلا أن المقابلة تتطلب وجود مقابلين مدربين على اجراءها. إذ لابد من مقابلة عدد كبير نسبياً من الأفراد.

3. الاستبانة Questionnaire:

أداة الاستبانة هي الأداة الأكثر استخداماً في البحوث الإنسانية والاجتماعية وهي تعتبر وسيلة لجمع البيانات من خلال احتوائها على مجموعة من الأسئلة أو العبارات و الطلب من المبحوثين الإجابة عنها.

ب-المصدر الثانويةSecondary Source: وتعتمد هذه المصادر في جمع البيانات على جهات قامت بجمع البيانات باستخدام الطرق الأولية وتعارف المصادر الثانوية باليد الثانية للتحليل **Second-Hand Analysis**لكون الباحث يقوم على تحليل البيانات الموجودة مسبقاً بشكل مختلف أو يسعى الى الإجابة على تساؤلات أخرى. ويمكن أن تكون المصادر الثانوية إما كمية مثل التقارير الصادرة عن دائرة الإحصاءات، أو نوعية مثل الكتب و الأبحاث.

9. تصميم الاستبانة Questionnaire Design:

يقوم الباحث بتصميم الاستبانة وتوزيعها على الأفراد المبحوثين، ويتم تقسيم عملية تصميم الاستبانة الى ثلاث أقسام رئيسه :

- ✓ **القسم الأول** : يتعلق بالمقدمة أو ما يسمى برسالة التغطية **CoverLetter** حيث يتم في هذا القسم التعريف بعنوان البحث بالإضافة الى التأكيد على سرية المعلومات التي سيتم جمعها من المبحوثين.
- ✓ **القسم الثاني** :يتناول ارشادات وتعليمات تعبئة الاستبانة وكيفية الإجابة على أسئلتها.
- ✓ **القسم الثالث** : يحتوي على العبارات والأسئلة المتعلقة بمتغيرات البحث موضوع الاستبانة.

ويمكن ايجاز خطوات تصميم الاستبانة في الآتي :

1- **تحديد أهداف البحث** : الخطوة الأولى في تصميم الاستبانة هي تحديد الأهداف التي من أجلها ستصمم هذه الاستبانة مع ضرورة التأكد من ترجمة تلك الأهداف الى أسئلة أو عبارات لأجل الحصول على البيانات اللازمة.

2- **اقرار طريقة تعبئة البيانات اللازمة** : إقرار شكل البيانات **Format** من حيث طريقة الإجابة المفتوحة أو المغلقة واللغة المستخدمة.

3- **صياغة الأسئلة** : البدء بالسؤال الذي يثير اهتمام المبحوث. ومراعاة وضع الأسئلة العامة أولاً ثم الانتقال الى الأسئلة التي تحتاج الى وقت وجهد وارجاء الأسئلة التي يتوقع أن تكون مثار جدل الى نهاية الاستبانة.

وهناك اجمالاً نوعان من أسئلة الاستبانة :

أ- **الأسئلة المفتوحة الاجابة** :تلك الأسئلة تكون اجاباتها مفتوحة مثل ما رأيك في......وضح الأسباب التي........وهكذا. ويستخدم هذا النوع من الأسئلة بكثرة في البحوث الاستكشافية محاولة من الباحث للحصول على أكبر قدر ممكن من البيانات.

ب-**الأسئلة المغلقة الاجابة** : تكون اجابات تلك الأسئلة محددة بخيارات معينة، وما على المبحوث إلا أن يقوم بالتأشير على الاجابة التي يختارها، وقد تكون الخيارات ثنائية أي نعم/لا. أو قد تكون متعددة مثل أوافق بشدة / أوافق / غير متأكد / غير مرافق / غير موافق اطلاقاً. كما قد تستخدم في جمع بيانات رقمية عن الوحدة المبحوثة كالعمر والجنس ومستوى التعليم.

4- **تقييم الاستبانة** : هل الأسئلة أطول من اللازم، وهل تزودنا اجابات الاستبانة بالمعلومات التي تحقق الأهداف منها.

5- **تحديد طريقة توزيع الاستبانة** : هل سيتم توزيعها باليد أو بالبريد أو بأية وسيلة أخرى.

6- **اجراء الاختبار القبلي** : اختبار عينة صغيرة لإجراء اختبار الاستبانة من خلالها وذلك للتأكد من سهولة أو صعوبة الأسئلة حيث يتم الاهتمام بآراء أفراد هذه العينة واجراء التعديلات اللازمة على الاستبانة تبعاً لمنطقية ومعقولية آرائهم.

7- **الشكل النهائي للاستبانة** : ويتم اعداد الشكل النهائي للاستبانة وطباعتها مع تدقيق الطباعة للتأكد من خلوها من الأخطاء المطبعية.

بعد عملية تصميم الاستبانة فإنه يتم توزيعها على الجهة المستهدفة وجمعها بعد تعبئتها تمهيداً لإجراء التحليلات الاحصائية اللازمة عليها.

10.ترميز البيانات:

بعد القيام بجمع البيانات و تصنيفها يقوم المحلل الإحصائي بترميز تلك البيانات تمهيدا لإدخالها الى البرنامج، إن المقصود بترميز البيانات هو اعطاء كل اجابة من اجابات المبحوث رقماً معيناً أو حرفاً محدداً لكي تتم عملية ادخال هذه الاجابات الى البرنامج بسهولة وبشكل يكون البرنامج قادراً على التعامل معها.

وفيما يلي سنورد بعض الأمثلة في هذا المجال :

الأسئلة التي تتطلب اجابتها خيارين:

(ذكر) رقم (2)   (أنثى) رقم (1)

(نعم) رقم (2)   (لا) رقـــم (1)

الأسئلة التي تحتمل أكثر من اجابتين :

(أوافق بشدة) رقم (5) (دائما) رقم (5) (راضي جداً) رقم (4) (أوافق) رقم (4)

(غالبا) رقم (4) (راضي) رقم (3)

(غير متأكد) رقم (3) (أحياناً) رقم (3) (غير راضي) رقم (2)

(لا أوافق) رقم (2) (نادراً) رقم (2)   (غير راضي أبداً) رقم (1)

(لا أوافق أبداً) رقم (1) (أبداً) رقم (1)

وبالتالي فقد تشتمل الاجابات على خيارين(2.1) أو ثلاثة خيارات (3.2.1) أو أربعة خيارات أو خمسة خيارات (5.4.3.2.1) أو اكثر من ذلك عندما يقوم الباحث بطباعة رقم (5) بدلاً من لأوافق بشدة ورقم (4) بدلاً من أوافق. فإن البرنامج في هذه الحالة سوف يتعامل مع أرقام أو قيم ورموز وبالتالي يمكن استخراج العشرات من الاحصاءات المعروفة كالوسط الحسابي والانحراف المعياري واختيار (ت) واختيار (ف) ومعامل الارتباط والانحدار وغير ذلكمن الاحصاءات التي يحتاجها الباحث.

11. صياغة واختيار الفرضيات:

الفرضيات هي عبارة عن حلول للمشكلة قيد الدراسة سوف يثبت صحتها أو عدم صحتها بعد الانتهاء من تحليلها و دراستها. وهناك فرق جوهري بين الفرضيات **Hypothesis** و بين الافتراضات **Assumption** ففرضيات البحث هي عبارة عن اجابات محتملة لأسئلة البحث مستمدة من خلفية علمية. ويمكن التحقق من قبولها أو رفضها من خلال المعلومات التي تجمع عنها وتحليلها. أما الافتراضات فالمقصود بها مسلمات البحث أي ما ينبغي أن يتم التسليم بصحتها حيث أنها لا تتعارض مع الحقائق العلمية ولا تحتاج الى ادلة و براهين لإثبات صحتها.

- **صياغة الفرضيات :**

يقوم الباحث ببناء الفرضيات المتعلقة بدراسته، حيث يعتمد في بناء الفرضيات على أسس معينة مثل المنطق أو الملاحظات الشخصية أو قد يعتمد على توقع وجود علاقة معينة بين متغيرين يفترض هو شخصياً وجود علاقات بينهما.

وهناك عدة طرق لصياغة الفرضيات من أهمها :

أ- الطريقة المباشرة :

وفقاً لهذه الطريقة يصوغ الباحث فرضياته بطريقة مباشرة أي بصيغة الاثبات وهنا يتوقع الباحث بدرجة كبيرة صحة الفرضية التي يضعها. وبالتالي فإنه يصوغ الفرضية ويبدأ بالبحث عن الأدلة والبراهين التي تدعم قبول الفرضية، أو تثبت عدم قبولها.

ب- الطريقة الإحصائية :

في حالة استخدام الباحث الأساليب الكمية لاختيار فرضيات البحث فإنه يقوم بإتباع الطريقة الاحصائية في صياغة الفرضيات. وحسب الطريقة الاحصائية يضع الباحث الفرضية الصفرية أو العدمية **NullHypothesis** أو باختصار **(Ho)** والتي تنص على عدم وجود تأثير أو عدم وجود علاقة بين المتغيرين قيد البحث. بالإضافة الى ذلك تكون هناك الفرضية البديلة **Alterative Hypothesis** أو **H_1** والتي تنص على وجود تأثير أو وجود علاقة بين المتغيرين قيد البحث.

فإذا كانت نتيجة التحليل الاحصائي قبول الفرضية الصفرية فإن الباحث يصل الى نتيجة أنه لا يوجد علاقة بين المتغيرين أما إذا كانت نتيجة التحليل الاحصائي رفض الفرضية الصفرية. فمعنى ذلك أن الباحث يقرر قبول الفرضية البديلة. والتي تنص على وجود علاقة بين المتغيرين.

أما من حيث تحديد اتجاهات الفرضية، فقد تكون الفرضية :

أ- غير محدد الاتجاه **Non-Direction al Hypothesis**

قد لا يملك الباحث توقعات وشواهد للتأكد من اتجاه الفرضية، وبالتالي فإنه يجعلها غير محددة الاتجاه.

....لا يوجد علاقة ذات دلالة احصائية بين التدريب وأداء العاملين.

$\mu \neq 30$...

وفي هذه الحالة نجري احصائي الاختبار ذو الطرقين **Two-Tailed**

ب-محددة الاتجاه **Direction al Hypothesis**

قديكون لدى الباحث أسباباً وشواهد للتأكد من اتجاه الفرضية فيجعل الفرضية محددة الاتجاه.

...لا يوجد علاقة طردية ذات دلالة احصائية بين الأسلوب الديكتاتوري في الإدارة وبين غياب العاملين.

$\mu > 30$...

وفي هذه الحالة نجري احصائي الاختبار ذو الطرف الواحد **One-Tailed**

➢ **اختبار الفرضيات :**

إذا قام الباحث بسحب عينة عشوائية من ومجتمع ما فإنه قد يكون هناك فرق بين وسط العينة ووسط المجتمع الأصلي. هذا الفرق يسميه الاحصائيون خطأ الصدفة أو خطأ المعاينة **SimplingError**.حيث يعتبر فرقاً غير معنوي، وبناء عليه، قد يستخدم الوسط الحسابي للعينة لتقدير **Estimate**الوسد الحسابي للمجتمع غير المعلوم.

أما إذا قام الباحث بسحب عينة من مجتمع معين وقام بمقارنة وسط هذه العينة مع وسط مجتمع آخر (معلوم)غير الذي تم سحب العينة منه فإن الفرق قد يكون ذا دلالة احصائية أو بمعنى آخر فرقاً معنوياً.

ويقول رجاء محمود أبو علاء **(2003)** بأنه لاتخاذ القرار حول الفروق بين عينتين عما إذا كانت فروقاً حقيقية وأن العينتين تنتميان لمجتمعين مختلفين لابد من تطبيق اختبار الدلالة الاحصائية. فاختبارات الدلالة الاحصائية تمكنك من تقدير احتمال أن البيانات الواردة من مجموعتين منفصلتين هما في الواقع تنتميان الى مجتمع واحد. وإذا لم يكن من المحتمل انهما اتيتا من مجتمع واحد. يمكنك أن تتخذ قراراً بذلك، وإذا كان من المحتمل أن الفروق بين مجموعتين (عينتين) وليده التغيرات الراجعة

الى الصدفة. نقول بأنه لا توجد فروق دالة بينهما، أما إذا كانت الفروق لا ترجع الى الصدفة، فإننا نقول أن من المحتمل أن هناك فروقاً حقيقية دالة احصائياً بين المجموعتين (العينتين).

ويؤكد مجدي عبد الكريم حبيب (2001)، أنه يمكن أن يقع الباحث في واحد من نوعين من الخطأ:

الخطأ من النوع الأول (Type 1 error) و الخطأ من النوع الثاني (Type 2 error). ويعرف النوع الأول من الخطأ بأنه رفض الفرضية الصفرية عندما تكون هذه الفرضية في الواقع صحيحة. أما النوع الثاني من الخطأ فهو قبول الفرضية الصفرية عندما تكون الفرضية في الواقع غير صحيحة.

وعادة ما يرمز الى احتمال وقوع الباحث في الخطأ من النوع الأول بالرمز ($\propto$)(ألفا باللغة اللاتينية). أما احتمال وقوع الباحث في الخطأ من النوع الثاني فيرمز له بالرمز β(بيتا باللغة اللاتينية) ويتم تحديد قيمة ($\propto$)من قبل الباحث (01/0.05/0.10/0.0) حسب طبيعة الدراسة التي يجريها ووفقاً لدرجة الثقة المطلوب وجودها في نتائج البحث.

عندما يقوم الباحث باختبار أي فرضية فهو في النهاية وبعد إجراء التحليل اللازم لها إما أن يقبل الفرضية أو يرفضها. ويتبع الباحث إجمالاً أسلوبينلاتخاذ قرار قبول أو رفض الفرضية الصفرية.

1- المقارنة مع القيمة الجدولية :

يتم احتساب قيمة إحصائي الاختبار **Test Statistic** ومقارنتها مع القيمة الحرجة التي تأخذ مساحة ($\propto$)والتي تستخرج من جدول التوزيع المتعلقة بإحصائي الاختبار فإذا كانت القيمة المحسوبة أقل من القيمة الجدولية فإننا نقبل الفرضية الصفرية. أما إذا كانت القيمة المحسوبة أكبر من القيمة الجدولية نرفض الفرضية الصفرية و نقبل البديلة.

2- استخدام P-Value :

هماك طريقة أخرى للرفض و القبول تعتمد على احتمال محسوب يسمى **P-Value** ويرمز له في **SPSS**بالرمز **Sig** ويعرف بأنه مستوى معنوية محسوب أو خطأ من النوع الأول محسوب.

ويمكن تعريف **P-Value**بأنها نسبة احتمال الحصول على قيمة متطرفة لإحصائي اختبار أكبر من القيمة المشاهدة بالصدفة وذلك في حالة كون الفرضية الصفرية صحيحة. وبالتالي فإن قاعدة القرار تشير الى قبول الفرضية الصفرية إذا كانت نسبة **P-Value** أكبر من مستوى المعنوية

المحدد سلفاً من قبل الباحث (**0.01** أو **0.05** أو **0.01**). أما إذا كانت نسبة **P-Value** تساوي أو أقل من مستوى المعنوية المحدد فإن قاعدة القرار تنص على رفض الفرضية الصفرية وبالتالي قبول الفرضية البديلة.

تحليل التباين Analysis Of Variance ANOVA :

مقدمة :

يعد أسلوب تحليل التباين (ANOVA)أحد الطرق أو الأساليب الاحصائية التي تستخدم لقياس الفروق بين متوسطات (Means) عدد من المجتمعات. عندمايصعب أو يتعذر علينا قياسها باستخدام اختبار (T)الذي يفضل استخدامه في حالة قياس الفرق بين متوسطين أو ثلاثة متوسطات فقد. وفي الحالة التي يكون فيها عدد المتوسطات أكثر من ذلك، فإن أنسب طريقة لقياس الفروق فيما بينها، هو اختبار (F)الذي يمكن الحصول عليه من خلال جدول تحليل التباين(ANOVA)، ويكون أسلوب تحليل التباين على عدة أنواع، نذكر منها:

1) تحليل التباين باتجاه واحد **One – wayAnalysis Of Variance**
2) تحليل التباين باتجاهين **Two – wayAnalysis Of Variance**

وفيما يلي شرحاً مفصلاً لكلا النوعين/ وعلى النحو الآتي :

1. تحليل التباين باتجاه واحد One wayANOVA:

يستخدم أسلوب تحليل التباين باتجاه واحد (أحادي التصنيف)، عندما يكون المتغير المستقل(X_i) منالنوع المصنف(Categorical Variable) ويمثل عدد المجتمعات (المجموعات)، أما المتغير المعتمد(Y_{ij}) فإنه يمثل المشاهدات (Observations)، ويخضع هذا المتغير للتوزيع الطبيعي (Normal Distribution)، ويفترض هذا الأسلوب أن تكون المجتمعات مستقلة بعضها عن البعض الآخر.

بافتراض لدينا (K) من المجتمعات، اختيرت عيناتها بشكل عشوائي حجم كل منها(n)من المشاهدات وبفرض إن هذه المجتمعات مستقلة بعضها عن البعض وتخضع للتوزيع الطبيعي بمتوسط[μ_1,μ_2, ...μ_k]،وبتباينات متساوية $(\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2)$

والمطلوب هو اختيار الفروق بين متوسطات المجتمعات (المجموعات).

والجدول التالي: يوضح توزيع المشاهدات (Y_{ij})رقم(j)المأخوذة من المجتمع(i) :

المجتمعات Populations	المشاهدات (Y_{ij})					Total	المتوسطات Means
	1	2	j	....	n		
1	Y_{11}	Y_{21}.. . .	Y_{ij}	....	Y_{1n}	$Y_{1.}$	$\bar{Y}_{1.}$
2	Y_{21}	Y_{22}	Y_{2j}	....	Y_{2n}	$Y_{2.}$	$\bar{Y}_{2.}$
.	.		.		.	.	.
.	.		.		.	.	.
.	.		.		.	.	.
I	Y_{i1}	Y_{i2}	Y_{ij}	....	Y_{in}	$Y_{i.}$	$\bar{Y}_{i.}$
.	.		.		.		.
.	.		.		.		.
.	.		.		.		$\bar{Y}_{k.}$
K	Y_{k1}	Y_{k2}	Y_{kj}	....	Y_{kn}	$Y_{k.}$	
Total	– .... –	..............–	....		–	$Y_{..}$	$\bar{Y}_{..}$

حيث أن :

$Y_{i.}\sum y_i$ تمثل مجموع المشاهدات للمجتمع رقم **(i)**

$\bar{Y}_{i.}$: تمثل متوسط المشاهدات للمجتمع رقم**(i)**

$Y_{..}\sum y_{ij}$ تمثل مجموع المشاهدات الكلية

$\bar{Y}_{..}$: تمثل المتوسط العام للمشاهدات الكلية.

وبناءاً على ما تقدم، فإن المشاهدة (y_{ij})، يمكن التعبير عنها بالنموذج الآتي :

$$Y_{ij} = \mu_i + \epsilon_{ij} \qquad \text{.................}(1)$$

حيث أن :

$Y_{i.}$:تمثل قيمة المشاهدة **(j)**المأخوذة من المجتمع رقم **(i)**

μ_i:تمثل متوسط المجتمع رقم**(i)**

$\in_{ij}$:مقياسانحراف المشاهدة (j)في المجتمع (i)عم متوسط المجتمع (μ_i)

إن الفرضية الاحصائية المكافئة للنموذج رقم (1)، هي :

$$H_0 = \mu_1 = \mu_2 = \ldots\ldots = \mu_k$$

H_1 **= At Least twomeans are not equal**

يمكن إعادة كتابة المعادلة رقم (1)، بعد التعويض عن (μ_i)بما يساويها، إذ إن :

$$\mu_i = \mu + \propto_i$$

عليه يكون النموذج البديل للمشاهدة **(Y_{ij})**، على النحو الآتي :

$$\boldsymbol{Y_{ij} = U_i + X_j + E_{ij}}\text{................ (2)}$$

(i= 1,2,……,k, j=1,2, ….,n)

تحت القيد$(\sum_i^k \propto_i = 0)$.

حيث إن:

μ:يمثل متوسط العام لمتوسطات المجتمعات (μ_i).

$\propto_i$:تمثل تأثير المجتمع رقم **(i)**.

وبالتالي تكون الفرضية الاحصائية المكافئة للنموذج رقم (2)، تأخذ الشكل الآتي :

$$H_0 = \propto_1 = \propto_2 = \ldots\ldots = \propto_k = 0$$

H_1 = At Least oneof ($\propto_i$is not equal to (zero).

وتأسيساً على ما تقدم، فإن اسلوب تحليل تباين باتجاه واحد، ينطوي على تجزئة مجموع المربعات الكلي الى مركبتين تتمثل بـ [مجموع مربعات يُعزى للتفاوت بين المعاملات (المجموعات)، و الثانية تمثل مجموع مربعات يُعزى للخطأ التجريبي]، أي أن :

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

$$=n\sum_{i}^{k}(\bar{Y}_{i.}-\bar{Y}_{..})^2+\sum_{i}^{k}\sum_{j}^{n}(Y_{ij}-\bar{Y}i_{.})^2 \ldots\ldots (3)$$

إن مكونات العلاقة رقم **(3)**، يمكن التعبير عنها بدلالة مجموع المربعات**(Sum of Squares)**أي إن :

SST = SSG + SSE............ (4)

SST : تمثل مجموع المربعات الكلي **(Total Sum of Squares)**.

SSG : تمثل مجموع المربعات المجموعات**(Groups Sum of Squares)**.

SSE : تمثل مجموع المربعات الخطأ**(Error Sum of Squares)**.

ولأغراض الحساب، وبناء جدول تحليل التباين **(ANOVA)**، يمكن استخدام المعادلات التالية للحصول على مجموع المربعات الخاصة بـ **(SSE , SSG , SST)**على النحو الآتي :

$$\text{SST}=\sum_{i}^{k}\sum_{j}^{n}Y_{ij}^2-nk\,\bar{Y}^2 \ldots\ldots (5)$$

$$\text{SSG}=n\sum_{i}^{k}\bar{Y}_{i.}^2-nk\,\bar{Y}^2 \ldots\ldots (6)$$

$$\text{SSE}=\text{SST}-\text{SSG} \ldots\ldots (7)$$

والجدول التالي : يوضح مكونات تحليل التباين باتجاه واحد.

مصدر الاختلاف **Source Of Variation**	مجموع المربعات **Sum Of Squares**	دراجات الحرية **Degree Of Freedom**	متوسط المربعات **Mean Of Squares**	قيمة(F) المحسوبة**Fractio**
بين المجموعات B. Groups	SSG	K – 1	$MSG=\frac{SSG}{k-1}$	$F=\frac{MSG}{MSE}$
Error الخطأ	SSE	K(n – 1)	$MSE=\frac{SSE}{k(n-1)}$	
Total الكلي	**SST**	**Nk - 1**	–	–

بعد ذلك يتم مقارنة قيمة (F)المحسوبة، مع قيمة(F)الجدولية التي يتم الحصول عليها من جداول توزيع(F)، بدرجتي حرية البسط (k − 1) و المقام [k (n − 1)]، و مستوى معنوية معين (∝)، أي أن

$$\left\{\left(F_{[\![(k-1),k(n-1)\propto]\!],}\right)\right\}$$

قاعدة القرار :

إذا كانت قيمة (F)المستخرجة أكبر من أو تساوي قيمة (F)الجدولية، يدل ذلك على رفض الفرضية (H_0)، وهذا يعني وجود فروق معنوية بين المجموعات، عند مستوى المعنوية(∝)، وبعكسه يتم قبول الفرضية(H_0)، مما يعني عدم وجود فروق بين المجموعات.

مثال (1) :

البيانات الموضحة بالجدول التالي، تمثل نتائج تجربة زراعية لثلاثة أنواع من السماد الكيمياوي المركب، ثم استخدامها لزيادة إنتاجية محصول الحنطة، لعدد من الأراضي الزراعية بنفس المساحة والخصوبة.

أنواع السماد الكيمياوي	المشاهدات (Y_{ij})								Total
	1	2	3	4	5	6	7	8	
t_1	8	10	7	6	9	5	8	7	**60**
t_2	5	3	5	4	4	6	5	4	**36**
t_3	15	10	11	13	9	13	14	11	**96**
Total	–	–	–	–	–	–	–	–	192

المطلوب :

1- صياغة الفرضية الاحصائية للتجربة اعلاه.

2- استخدام أسلوب تحليل التباين **(ANOVA)**، لاختيار الفروق بين متوسطات أنواع السماد الكيمياوي، مستخدما مستوى المعنوية $(\propto = 0.01)$.

الحل :

1− الفرضية الاحصائية للتجربة هي :

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

H_0 **: At Least twomeans are not equal.**

2− لبناء جدول تحليل التباين باتجاه واحد **ANOVA(One −Way)**،نقوم بحساب المتوسطات$(\overline{Y}_{i.})$، و مجاميع المربعات **(Sum Of Squares)**،كالآتي :

$$\overline{Y}_{1.} = \frac{60}{8} = 7.5$$

$$\overline{Y}_{2.} = \frac{36}{8} = 4.5$$

$$\overline{Y}_{3.} = \frac{96}{8} = 12.0$$

$$\therefore \overline{Y}_{..} = \frac{7.5 + 4.5 + 12}{3} = 12.0$$

$$= 8$$

أو يمكن إيجاد المتوسط العام $(\overline{Y}_{..})$،وفقاً للصيغة الآتية :

$$\frac{\overline{Y}_{..} = \sum_{i}^{3}\sum_{j}^{8} Y_{ij}}{nk}$$

$$= \frac{192}{8(3)}$$

$$=8$$

$$\therefore SST = \sum_{i}^{3}\sum_{j}^{8} Y_{ij}^2 - nk\,\overline{Y}^2$$

$$= [8^{/2} + 10^{/2} + \cdots + 11^{/2}] - 8(3)(8)^2$$

$$= 1818 - 1536$$

$$= 282$$

$$SSG = n\sum_{i}^{3} \overline{Y}_i^2 - nk\,\overline{Y}^2$$

$$= 8\,[7.5^{/2} + 4.5^{/2} + 12^{/2}] - 8(3)(8)^2$$

$$=1764 - 1536$$

$$=228$$

$$\therefore SSE = SST - SSG$$

$$= 282 - 228$$

$$=54$$

عليه يكون جدول تحليل التباين باتجاه واحد **(ANOVA(One – Way**،كالآتي :

مصدر الاختلاف **S.O.V**	مجموع المربعات **SS**	درجات الحرية **Df.**	متوسط المربعات **MS**	قيمة(F) المحسوبة **Fractio**
بين المعاملات	228	2	114	**F = 44.333****
الخطأ	54	21	2.5714	
الكلي	282	23	–	–

تقوم باستخراج قيمة **(F) الجدولية، من جدول توزيع (F)**، بدرجتي حرية **(21،2)**، و مستوى المعنوية $(\propto = 0.01)$ **تبين بأن** $F_{[(2.21;0.01)=5.78]}$

القرار :

يتضح من جدول تحليل التباين، بأن قيمة **(F)**المحسوبة بلغت **(44.333)** وهي أكبر من قيمة **(F)**الجدولية البالغة(5.78)،عند مستوى المعنوية $(\propto = 0.01)$، مما يدلل على رفض الفرضية H_0

وهذا يعني وجود فروق معنوية بين متوسطات أنواع السماد الكيمياوي (μ_1, μ_2, μ_3) عند مستوى المعنوية $(\propto = 0.01)$.

تحليل التباين باتجاهين Two – wayAnalysis Of Varieance :

عند دراستنا تحليل التباين باتجاه واحد، افترضنا وجود متغير أو مؤشر واحد على المتغير (Y_{ij})، وإن ما يتبقى من اختلاف أو تأثير بعد حساب أثر المدروس، فإنه يعزي الى العوامل العشوائية أو الصدفة. ولكن الحالة تختلف عند دراسة اسلوب التباين باتجاهين، فإن هذا الأسلوب يفترض دراسة تأثير متغيرين، أحدهما يمثل الصفوف **(Rows)**، أما الآخر فيمثل الأعمدة **(Columns)** على المتغير المعتمد (Y_{ij}) ، مثال ذلك دراسة تأثير المستوى العلاجي للمرضى وأساليب المعالج المتبعة على متوسط نسبة الشفاء من المرض (Y_{ij}).

ويعد أسلوب تحليل التباين باتجاهين، من أكثر الأساليب استخداما على مستوى الدراسات التربوية وعلم النفس والاجتماع، فعلى سبيل المثال يستخدم هذا الأسلوب لدراسة تأثير طرق التدريس ومستويات الطلبة على تحصيلهم العلمي، بالإضافة الى استخداماته في مجال الزراعة والصناعة والطب والإدارة وغيرها من العلوم الأخرى.

والجدول التالي، يوضح توزيع المشاهدات بواقع مشاهدة واحدة لكل خلية، إذ تشير المشاهدة (Y_{ij}) في العمود (j)المخصصة للصف رقم(i).

حيث أن :

$Y_{i.}$:تمثل متوسط المشاهدات الصفرقم**(i)**

$\overline{Y}_{i.}$:تمثل مجموع المشاهدات الصفرقم**(i)**

$Y_{.j}$:تمثل مجموع المشاهدات العمود رقم (j)

$\overline{Y}_{.j}$:تمثل المتوسط مشاهدات العمود رقم (j).

$Y_{..}$:تمثل مجموع المشاهدات الكلية.

الصفوف Rows	الأعمدة(Columnbs) 1 2 .. .. j c	Total	Means
1	Y_{11} $Y_{21..}$ . . Y_{ij} .... Y_{1c}	Y1.	$\overline{\mathbf{Y}}_{1.}$
2	Y_{21} Y_{22} .. . . Y_{2j} .... Y_{2c}	$Y_{2.}$	$\overline{\mathbf{Y}}_{2.}$
.	. . .	.	.
.	. . .	.	.
.	. . .	.	.
i	Y_{i1} $Y_{i2..}$ . . Y_{ij} .... Y_{ic}	$Y_{i.}$	$\overline{\mathbf{Y}}_{i.}$
.	. . .		.
.	. . .		.
.	. . .		.
r	Y_{r1} Y_{r2}. . . Y_{rj} ....Y_{rc}	$Y_{r.}$	$\overline{Y}_{r.}$
Total	$Y_{.1}$ $Y_{.2}$ ... Y_{ij}... $Y_{.c}$	$Y_{..}$	..
Means	$\overline{Y}_{.1}\overline{Y}_{.2}$ ... $\overline{Y}_{.j}$ .... $\overline{Y}_{.c}$	–	$\overline{\mathbf{Y}}_{..}$

$Y_{..}$:تمثل مجموع المشاهدات الكلية.

$\overline{Y}_{..}$:تمثل المتوسط العام للمشاهدات الكلية.

وبناءاً على ما تقدم، يمكن تمثيل المشاهدة **(Y_{ij})**، بالنموذج الآتي :

$$Y_{ij} = \mu_i + \epsilon_{ij} \qquad ....(1)$$

حيث إن :

Y_{ij}:تمثل قيمة المشاهدة في العمود (j) المخصصة للصف رقم (i)

μ_{ij} : تمثل متوسط المجتمع.

$\in_{ij}$: تمثل مقياس انحراف المشاهدة $Y_{.j}$عن متوسط المجتمعμ_{ij}.

يمكن إعادة كتابة المعادلة رقم (1)، بعد التعويض عن (μ_{ij}) بما يساويها، إذ إن :

$$\mu_{ij} = \mu + \propto_i + \beta_j$$

عليه يكون النموذج البديل للمشاهدة (Y_{ij})، يعطى بالشكل الآتي :

$$Y_{ij} = \mu + \propto_i + \beta_j + \varepsilon_{ij} Y_{ij} \qquad ....(2)$$

$$(i = 1,2,.....,r \quad , \quad j = 1,2,.....,c)$$

تحت القيد

$$\left(\sum_{i}^{r} \propto_i = \sum_{j}^{c} \beta_j = 0 \right)$$

حيث إن :

μ:تمثل متوسط العام.

$\propto_i$: تمثل تأثير الصف رقم(i).

β_j : يمثل تأثير العمود رقم (j).

وتأسيساً على ما تقدم، تكون الفرضيات الاحصائية المكافئة للنموذج رقم **(2)**، على النحو الآتي :

(1)الفرضية الخاصة باختبار الفروق بين معدلات الصفوف **(Rows)**.

$$H_0 : \propto_1 = \propto_2 = \cdots ... = \propto_r = 0$$

H_1: At least one of ($\propto_i$) is not equal to (Zero).

وبناءاًعلى ما تقدم، فإن أسلوب تحليل التباين باتجاهين، ينطوي على تجزئة مجموع المربعات الكلي الى ثلاث مركبات تتمثل بـ [مجموع مربعات يعزي للتفاوت بين الصفوف، ومجموع مربعات يعزي للتفاوت بين الأعمدة، والثالثة تمثل مجموع مربعات تعزي للخطأ التجريبي]، أي إن :

$$\sum_{i}^{r}\sum_{j}^{c}(y_{ij}-\bar{y}_{..})^2=\sum_{i}^{r}\sum_{j}^{c}(\bar{Y}_{i.}-\bar{y}_{..})^2+\sum_{i}^{r}\sum_{j}^{c}(\bar{Y}_{.j}-\bar{y}_{..})^2+\sum_{i}^{r}\sum_{j}^{c}(\bar{Y}_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})^2$$

$$=c\sum_{i}^{r}(\bar{Y}_{i.}-\bar{Y}_{..})^2+r\sum_{j}^{c}(\bar{Y}_{.j}-\bar{Y}_{..})^2+\sum_{i}^{r}\sum_{j}^{c}(\bar{Y}_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})^2 \dots\dots\dots(3)$$

إن مكونات العلاقة رقم **(3)**، يمكن التعبير عناه بدلالة مجموع المربعات**(Sum Of Squares)**، أي إن :

SST = SSR + SSC + SSE..................... (4)

حيث إن :

SST : تمثل مجموع المربعات الكلي **(Total Sum of Squares)**.

SSR : تمثل مجموع المربعات الصفوف **(RowsSum of Squares)**.

SSC : تمثل مجموع مربعات الأعمدة **(ColumnsSum of Squares)**.

SSE : تمثل مجموع مربعات الخطأ **(ErrorSum of Squares)**.

ولأغراض الحساب، وبناء جدول تحلي التباين **(ANOVA)** يمكن استخدام المعادلات التالية للحصول على مجموع المربعات الخاصة بـ **(SSE, SSC, SSR, SST)**، و على المحو الآتي :

$$\mathbf{SST}=\sum_{i}^{r}\sum_{j}^{c}Y_{ij}^2-rc\,\bar{Y}^2 \dots\dots\dots(\mathbf{5})$$

$$\mathbf{SSR}=c\sum_{i}^{r}\bar{Y}_{i.}^2-rc\,\bar{Y}^2 \dots\dots\dots(\mathbf{6})$$

$$\mathbf{SSC}=r\sum_{j}^{C}\bar{Y}_{.j}^2-rc\,\bar{Y}^2 \dots\dots\dots.(\mathbf{7})$$

$$\mathbf{SSE}=SST-SSR-SSC \quad \dots\dots\dots\dots(\mathbf{8})$$

والجدول التالي، يوضح مكونات جدول تحليل التباين باتجاهين :

مصدر الاختلاف Source of variation	مجموع المربعات Sum of squares	درجات الحرية Degree of Freedom	متوسط المربعات Mean Of Squares	قيمة(F)المحسوبة Fractio
بينمعدلات الصفوف	SSR	r - 1	$MSR = \frac{MSR}{MSE}$	$F_r = \frac{MSR}{MSE}$
بين معدلات الأعمدة	SSC	c-1	$MSC = \frac{SSC}{c-1}$	$F_c = \frac{MSc}{MSE}$
الخطأ التجريبي	SSE	(r-1)(c-1)	$MSE = \frac{SSE}{(r-1)(c-1)}$	-
الكلي	SST	rc-1	-	-

بعد ذلك يتم مقارنة قيم **(F)**المحسوبة لكل من الصفوف و الأعمدة، مع قيمة**(F)**الجدولية التي يتم الحصول عليها من جدول توزيع**(F)**، بدرجتي حرية البسط و المقام، عند مستوى معنوية معين$(\propto)$.

قاعدة القرار:

1- إذا كانت قيمة**(F_r)**المحسوبة للصفوف أكبر من أو تساوي قيمة **(F)**الجدولية، يدلل ذلك على رفض فرضية العدم **(H_0)**، وهذا يعني وجود فروق معنوية بين معدلات الصفوف، عند مستوى المعنوية $(\propto)$، وبعكسه يتم قبول الفرضية **(H_0)**،مما يعني عدم وجود فروق بين معدلات الصفوف.

2- إذا كانت قيمة **(F_c)**المحسوب للأعمدة أكبر من أو تساوي قيمة**(F)**الجدولية، يدلل ذلك على رفض فرضية العدم **(H_0')**، وهذا يعني وجود فروق معنوية بين معدلات الأعمدة، عند مستوى المعنوية $(\propto)$، وبعكسه يتم قبول الفرضية **(H_0')**،مما يعني عدم وجود فروق بين معدلات الأعمدة.

مثال :

البيانات الموضحة بالجدول التالي، تمثل الإنتاجية لثلاثة أصناف من محصول القمح، عولجت الأراضي الزراعية بثلاثة أنواع من السماد الكيمياوي المركب، بافتراض بقاء العوامل الأخرى ثابتة.

أنواع السماد الكيمياوي المركب	أصناف القمح			Total
	w1	w2	w3	
t_1	95	100	105	**300**
t_3	70	110	60	**240**
t_1	60	105	105	**270**
Total	225	315	270	**810**

المطلوب :

1- صياغة الفرضيات الإحصائية للتجربة أعلاه.

2- استخدام أسلوب تحليل التباين باتجاهين، مستخدماً مستوى المعنوية (0.05)، لاحتبار الفروق.

أ- بين أنواع السماد الكيمياوي المركب.

ب-بين أصناف القمح.

الحل:

1-الفرضيات الاحصائية هي :

أ- الفرضية الاحصائية الخاصة باختبار الفروق بين أنواع السماد الكيمياوي المركب**(Rows):**

$$H_0 : \propto_1 = \propto_2 = \propto_3 = 0$$

H_1 : At least one of ($\propto_i$) in not equal to (Zero).

ب- الفرضية الاحصائية الخاصة باختبار الفروق بين أصناف القمح**(columns):**

$$H_0^{/} : \beta = \beta_2 = \beta_3 = 0$$

$H_1^{/}$: At least one of (β_j) in not equal to (Zero).

2- لبناء جدول تحليل التباين باتجاهين**(Two – way ANOVA)**، نقوم بحساب متوسطات الصفوف والأعمدة $(\overline{Y}_{i.})$ والأعمدة $(\overline{Y}_{.j})$، ومجاميع المربعات، على النحو الآتي :

$$\overline{Y}_{1.} = \frac{300}{3} = 100$$

$$\overline{Y}_{2.} = \frac{240}{3} = 80$$

$$\overline{Y}_{3.} = \frac{270}{3} = 90$$

ب-حساب متوسطات الأعمدة $(\overline{Y}_{.j})$:

$$\overline{Y}_{.1} = \frac{225}{3} = 75$$

$$\overline{Y}_{.2} = \frac{315}{3} = 105$$

$$\overline{Y}_{.3} = \frac{270}{3} = 90$$

عليه يكون المتوسط العام $(\overline{Y}_{..})$، كالآتي :

$$\overline{Y}_{..} = \frac{100+80+90}{3} = 90$$

أو يمكن ايجاد المتوسط العام $(\overline{Y}_{..})$، وفقاً للعلاقة الآتية :

$$\overline{Y}_{..} = \frac{\sum_{i}^{3}\sum_{j}^{3} Y_{ij}}{\text{rc}}$$

$$= \frac{810}{3(3)}$$

$$= 90$$

جـ. حساب مجموع المربعات $[SSE, SSC, SSR, SST]$، كالآتي :

$$SST = \sum_{\text{i}}^{3}\sum_{\text{j}}^{3} Y_{ij}^2 - rc\,\overline{Y}_{..}^2$$

$$[95^{/2} + \ 100^{/2} + \cdots + 105^{/2}] - 3(3)(90)^2$$

$$= 76300 - 72900$$

$$=3400$$

$$SSR = \text{c}\sum_{\text{i}}^{3} \overline{Y}_{i.}^2 - rc\,\overline{Y}_{..}^2$$

$$=3\left[100^{/2} + 80^{/2} + 90^{/2}\right] - 3(3)(90)^2$$

$$= 73500 - 72900$$

$$=600$$

$$SSC = r\sum_{j}^{3} \overline{Y}_{.j}^{2} - rc\,\overline{Y}_{..}^{2}$$

$$=3\left[75^{/2} + 105^{/2} + 90^{/2}\right] - 3(3)(90)^2$$

$$= 74250 - 72900$$

$$=1350$$

$$\therefore SSE = SST - SSR - SSC$$

$$= 3400 - 600 - 1350$$

$$=1450$$

عليه يكون جدول تحليل التباين باتجاهين **(Two – WayANOVA)**كالآتي :

مصدر الاختلاف **S.O.V**	مجموع المربعات **SS**	درجات الحرية **Df.**	متوسط المربعات **MS**	قيمة(F)المحسوبة **Fractio**
بينمعدلات الصفوف (بين أنواع السماد)	600	2	300	$F_r = 0.828$
بين معدلات الأعمدة (بين أصناف الحنطة)	1350	2	675	$F_c = 1.862$
الخطأ التجريبي	1450	4	362.5	–
الكلي	3400	8	–	–

من جداول توزيع **(F)**، وبدرجتي حرية البسط والمقام **(4,2)**، ومستوى المعنوية $(\propto = 0.05)$حصلنا على قيمة **(F)**الجدولية، $\left[F\left(_{2,4,0.05)=6.94}\right]\right.$.

القرار :

1−يتضح من جدول تحليل التباين، بأن قيمة(**Fr**)المحسوبة للصفوف، بلغت (**0.828**) وهي أقل من قيمة (**F**)الجدولية البالغة(**6.94**)، عند مستوى المعنوية (**0.05**)، مما يدلل ذلك على قبول فرضية العدم (H_0)، وهذا يعني عدم وجود فروق معنوية بين أنواع السماد الكيمياوي المركب، أي إن$[\propto_1 = \propto_2 = \propto_3 = 0]$

2−ويتضح أيضاُ، بأن قيمة (**Fc**) المحسوبة للأعمدة، بلغت (**1.862**) وهي أقل من قيمة(**F**)الجدولية البالغة (**9.64**)، عند مستوى المعنوية (**0.05**)، مما يدلل ذلك على قبول فرضية العدم

($H_0^{/}$) ، وهذا يعني عدم وجود فروق معنوية بين أصناف القمح، أي إن :

$[\beta_1 = \beta_2 = \beta_3 = 0]$.

تحليل الارتباط Analysis Of Correlation:

مقدمة :

يعد موضوع الارتباط من أكثر الطرق الاحصائية شيوعاً واستخداماً في دراسة العلاقة الارتباطية بين المتغيرات المدروسة لمختلفة العلوم (طبيعية كانت أم انسانية)، وتبرز الحاجة الى هذا النوع من التحليل في المجالات الاقتصادية والإدارية والتربوية، مثال ذلك :

دراسة العلاقة بين الدخل الشهري وحجم الإنفاق الأسري، أو دراسة العلاقة بين الإبداع في عناصر المزيج التسويقي والتفوق التنافسي لشركة ما، وغيرها من التغيرات لظواهر أخرى في الحياة العملية.

وسيتم التركيز في هذا الفصل على دراسة الارتباط الخطي البسيط والارتباط الجزئي والتعدد وكذلك اختبار معنوية العلاقة الارتباطية بين المتغيرات لكل نوع من أنواع الارتباط، كما سيتم تسليط الضوء بشيء من التفصيل على دراسة ارتباط الرتب.

1- مفهوم الارتباط:Concepet of Correlation

مما لا شك فيه إن أبرز ما يمكن الاهتمام به عند دراسة الارتباط بشكل عام، هو التعرف الى طبيعة واتجاه العلاقة الارتباطية ومقدار قوتها بين ظاهرتين أو أكثر، كالعلاقة بين المستوى الثقافي للأبوين والتحصيل الدراسي لأبنائهم، أو العلاقة بين متغير عدد أفراد الأسرة وحجم الإنفاق الشهري، أو العلاقة بين متغير مهارة العاملين وإنتاجية العامل اليومية في مصنع ما، وغيرها من الظواهر الأخرى في الحياة العملية.

وبناءاُ على ما تقدم، يمكن أن تكون العلاقة بين آية ظاهرتين ولتكن (X) و(Y)على عدة أشكال، فقد تكون العلاقة بين الظاهرتين طردية (موجبة)، بمعنى إن أية زيادة أو نقصان في قيمة المتغير(X) ستؤدي وبنفس الاتجاه الى زيادة أو نقصان في قيم المتغير (Y)، وقد تكون العلاقة عكسية (سالبة)، أي إن الزيادة في قيم المتغير(X) أو نقصانها، ستؤدي باتجاه معاكس الى نقصان في قيم المتغير (X) أو زيادتها.

عليه فإن الأساس في دراسة العلاقة الارتباطية بين المتغيرات، يستند الى العلاقة السببية التي تربط المتغيرات بعضها مع البعض الآخر، وهذا يعني وجود علاقة ارتباط منطقية تفسر سبب الارتباط بين المتغيرات، فعلى سبيل المثال لا الحصر، إن زيادة أوجه إنفاق أسرة ما على السلع والخدمات ناجم

عن عامل معين أو عدد من العوامل أدت الى هذه الزيادة، منها مثلاُ، ارتفاع الدخل الشهري للأسرة أو عدد أفراد الأسرة ...الخ، ففي هذا المثال تتضح العلاقة المنطقية بين المتغيرات إلا إنه ليس بالإمكانمن تحديد علاقة منطقية بين متغير مستوى ذكاء الطفل وحجم قدمه بالرغم من إمكانية حساب العلاقة الارتباطية بين المتغيرين، إلا إنه لا يمكن تفسير هذه العلاقة.

ويكون الارتباط الخطي على أنواع عدة، هي :

1. الارتباط الخطي البسيط. **Simple LinearCorrelation**
2. الارتباط الجزئي. **PartialCorrelation**
3. الارتباط المتعدد. **MultipleCorrelation**

وسيتم دراسة كل نوع من أنواع الارتباط السابقة بشيء من التفصيل، وعلى النحو الآتي :

2- الارتباط الخطي البسيط: SimpleLinearCorrelation

يعرف الارتباط الخطي البسيط بأنه "درجة العلاقة الارتباطية بين متغيرين فقط هما **(X)** و**(Y)**".

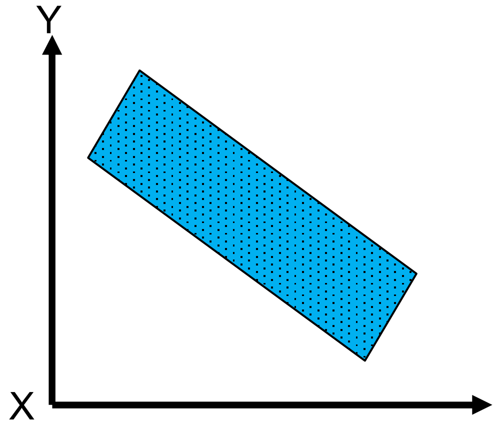
ويمكن قياس العلاقة الارتباطية بين المتغيرين على مرحلتين وكالآتي :

أ- الشكل الإنتشاري : ScatterDiagram

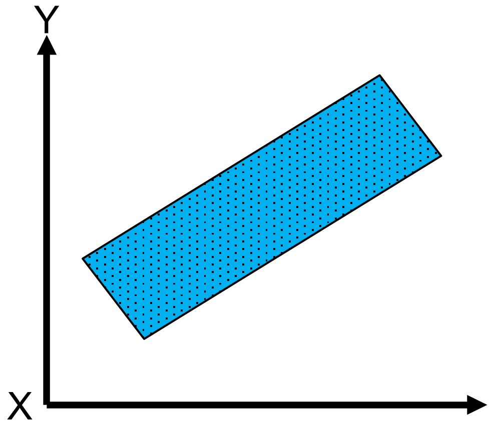
يعد الشكل الانتشاري من أبسط الطرق لعرض بيانات متغيرين يفترض بينهما علاقة ارتباطية، إذ يتم من خلاله تكوين فكرة أولية حول اتجاه العلاقة بين المتغيرين وقوتها.

وبالافتراض لدين متغيرين هما **(X)** و**(Y)**، على اساس عينة عشوائيةمن المشاهدات قوتها**(n)** فإن أزواج القيم لهذين المتغيرين تكتب على الشكل الآتي :

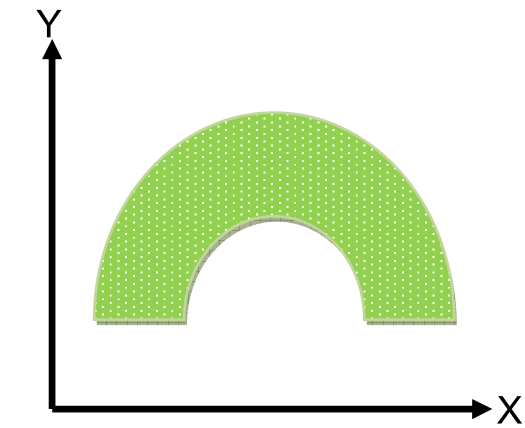
$[(X_n, Y_n), \dots, (X_2, Y_2), (X_1, Y_1)]$، وعند تمثيل أزواج قيم المتغيرين بواسطة الشكل الانتشاري، فإننا سنحصل على أحد الأشكال التالية، والتي من خلالها سيتم التعرف على العلاقة بين المتغيرين وقوتها.



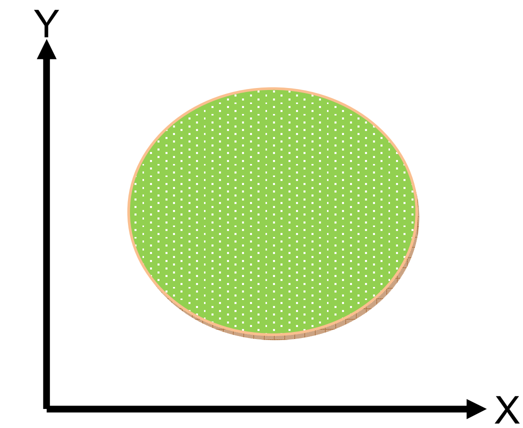
الارتباط سالب



الارتباط مرجب



ارتباط غير خطي (علاقة منحنية)



لا يوجد ارتباط بين المتغيرين

ب–معامل الارتباط البسيط: SimpleCorrelation Coefficient

يعرف معامل الارتباط البسيط بأنه : القيمة العددية للعلاقة الارتباطية الخطية بين المتغيرين ويأخذ معامل الارتباط البسيط عدد من الأشكال و الصيغ الرياضية، ويرمز له بالرمز **(r)**. وفيما يلي شرحاً مفصلاً لهذا المقياس :

- **معامل الارتباط البسيط لبيرسون :**

وهو مؤشر احصائي يستخدم لقياس القوة الارتباطية الخطية بين متغيرين كميين، أي (يمكن قياسهما كمياً)، مثال ذلك، قياس العلاقة بين الدخل الشهري للأسرة **(X)**وحجم أنفاقها الشهري**(Y)** ويعود الفضل الأول للعالم الإنجليزي كارل بيرسون**(Karl Pearson)** **(1867-1936)**في وضع الصيغة العامة لهذا المقاس.

ويمكن إيجاد معامل الارتباط البسيط، وفقاً لصيغة (بيرسون) الآتية :

$$r_p = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$$

حيث أن :

$$S_{xy} = \sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$$

$$S_{xx} = \sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$S_{yy} = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

3- خصائص معامل الارتباط البسيط بيرسون:

يتصف معامل الارتباط البسيط لبيرسون بالخصائص الآتية :

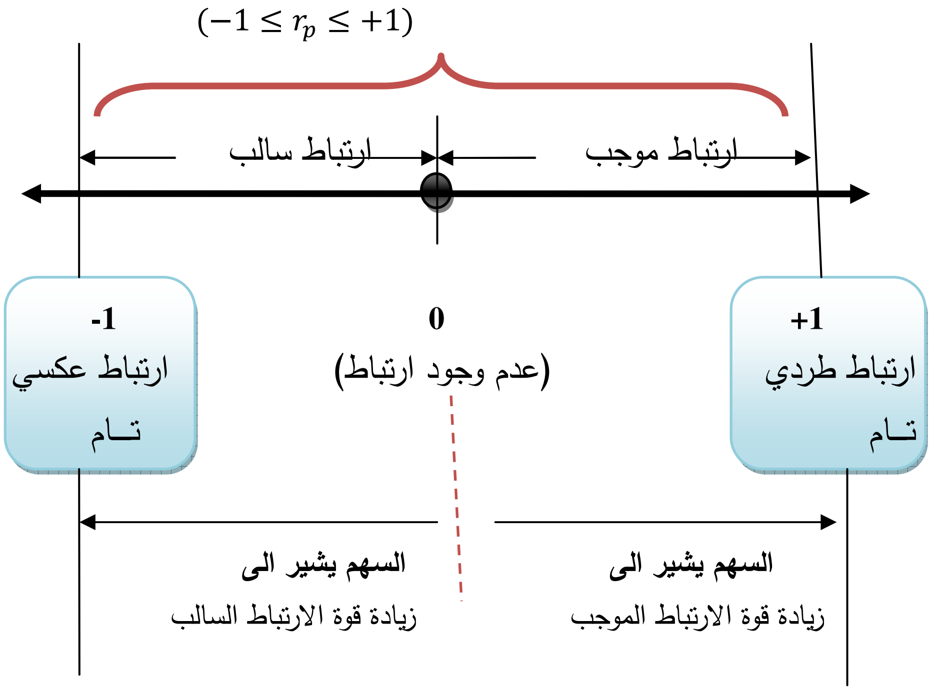
1- إن قيمة معامل الارتباط البسيط (r_p)تقع ضمن المجال $(-1 \leq r_p \leq +1)$، إذ إن:

أ) $r_p = +1$ ⟵ تعني إن الارتباط بين المتغيرين طردي وتام.

ب) $r_p = -1$ ⟵ تعني إن الارتباط بين المتغيرين عكسي وتام.

ت) $r_p = 0$ ⟵ تعني إن الارتباط بين المتغيرين معدوم.

والشكل التالي، يوضح ذلك :



2- إن أية تحويلات خطية تجرى على قيم أحد المتغيرين **(X)** أو **(Y)**، أو كليهما لا تؤثر على قيمة معامل الارتباط البسيط (r_p) المحسوب للقيم الجديدة.

3- تتأثر قيمة معامل الارتباط البسيط لبيرسون بالقيم الشاذة **(outlier Values)**،مما يتوجب على متخذ القرار، توخي الدقة عند تفسير قيمة هذا المعامل.

مثال (1):

وزن الطالب (x)	**75**	**72**	**69**	**67**	**63**
طول الطالب(Y)	172	170	172	165	162

البيانات التالية : تمثل أزواج القيم لأوزان (كغم) خمسة طلاب وأطوالهم (بالسم).

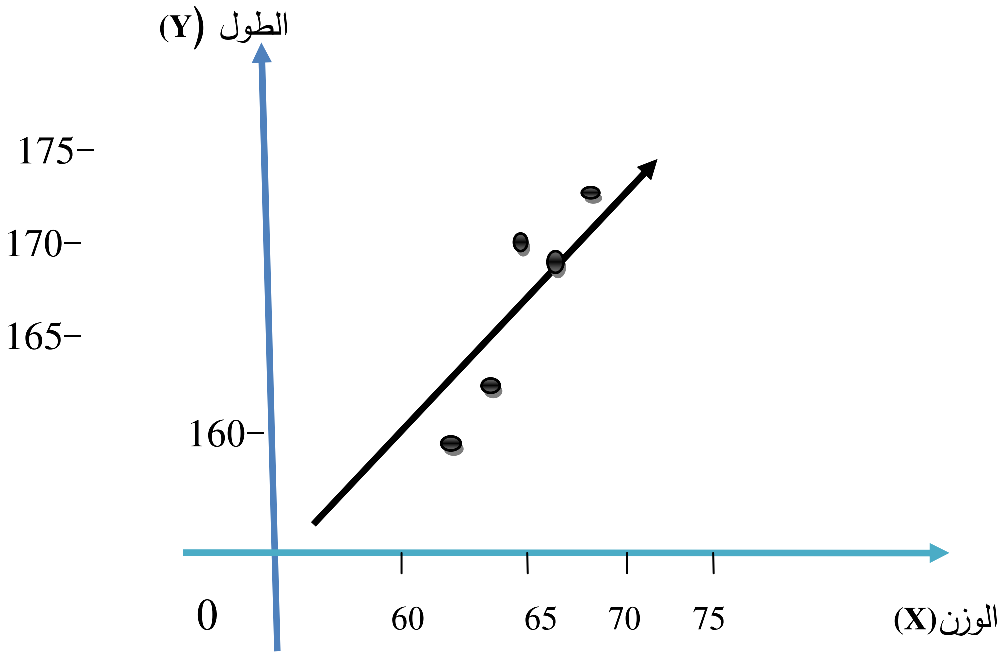
المطلوب :

1- ارسم الشكل الانتشاري **(Scatterdiagram)** لأزواج القيم.

2- احسب معامل الارتباط البسيط لبيرسون بين أوزان الطلبة وأطوالهم.

الحل :

1- رسم الشكل الانتشاري.



2- ايجاد معامل ارتباط بيرسون :

X	Y	$(X-\bar{X})^2$	$(X-\bar{X})^2$	$(Y-\bar{Y})(Y-\bar{Y})$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$
75	175	5.8	6.2	35.96	33.64	38.44
72	170	2.8	1.2	3.36	7.84	1.44
69	172	-0.2	3.2	-0.64	0.04	10.24
67	165	-2.2	-3.8	8.36	4.84	14.44
63	162	-6.2	-6.8	42.16	38.44	46.24
346	844	0	0	89.2	84.8	110.8
$\bar{Y} = 69.2\,,\bar{Y} = 168.8$				S_{xy}	S_{xx}	$\boldsymbol{S_{yy}}$

$$\therefore r_p = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$$

$$= \frac{89.2}{\sqrt{84.8}\sqrt{110.8}}$$

$$= \frac{89.2}{(9.21)(10.526)}$$

$$= \frac{89.2}{96.944}$$

$$\approx \ +0.92$$

يتضح من النتيجة أعلاه، بأن العلاقة بين وزن الطالب (X) و طوله (Y) علاقة طردية و قوية جدا. وهذا يعني بأن الزيادة في طول (Y)، تؤدي و بنفس الاتجاه الى زيادة وزنه (X).

3- اختبار معنوية معامل الارتباط البسيط لبيرسون :

بافتراض إن (r_p)يمثل معامل الارتباط البسيط لبيرسون بين أزواج قيم المتغيرين(**X**)و(**Y**)محسوب على أساس عينة عشوائية قوامها (n)فإن اختبار معنوية معامل الارتباط البسيط يكون على نوعين وكما يلي :

- **أولا** : اختبار معنوية معامل الارتباط عن قيمة مفترضة $(\rho = \rho_0)$:

لاختبار معنوية معامل الارتباط البسيط لبيرسون (r)في هذه الحالة، نقوم باختبار إحدى الفرضيات الاحصائية الثلاث، الآتية :

أ- الاختبار ذو جانبين : **Two – Tailed Test**

$$H_0: \rho = \rho_0$$
$$H_1: \rho \neq \rho_0$$

ب-الاختبار ذو جانب علوي :**Upper – Tailed Test**

$$H_0: \rho \leq= \rho_0$$
$$H_1: \rho > \rho_0$$

ت-الاختبار ذو جانب سفلي :**Lower – Tailed Test**

$$H_0: \rho \geq \rho_0$$
$$H_1: \rho < \rho_0$$

حيث إن:

ρ : تمثل معامل الارتباط البسيط بين المتغيرين (**X**)و(**Y**)في المجتمع.

ρ_0 : تمثل قيمة مفترضة غير مساوية للصفر.

إن إحصاءه الاختبار الملائمة للفرضيات الاحصائية السابقة، تأخذ الشكل الآتي :

$$z_{cal.} = \frac{w - E(w)}{\sqrt{Var(w)}} - N(0,1)$$

حيث إن :

$$\sqrt{E(w)=\frac{1}{2}ln\left(\frac{1+P_0}{1-P_0}\right)}$$

$$w=\frac{1}{2}Ln\left(\frac{1+r}{1-r}\right)$$

$$Var(w)=\frac{1}{n-3}$$

ولاتخاذ القرار الاحصائي حول رفض أو عدم رفض فرضية العدم (H_0)يتم مقارنة القيمة المطلقة لإحصاءة الاختبار المحسوبة $(|Z_{cat.}|)$ مع القيمة الجدولية $(Z_{tab.})$، إعتمادا على

مستوى المعنوية $(\propto)$، ونوع الفرضية البديلة (H_1)، و الجدول التالي، يوضح بعض القيم الجدولية الشائعة الاستخدام لـ (Z) التي تخضع للتوزيع الطبيعي المعياري [N(0,1)].

H_1	$H_1: \rho > \rho_0 , \rho < \rho_0$	$H_1: \rho \neq \rho_0$
$\propto$	$Z_{\propto}$	$Z_{\propto/2}$
10%	1.282	1.645
5%	1.645	1.960
1%	2.326	2.576

القرار الاحصائي :

أ- يتم قبول فرضية العدم (H_0)، عندما تكون القيمة المطلقة لإحصاءة الاختبار المحسوبة أقل من القيمة الجدولية، أي إن $[|Z_{cal.}| < Z_{tab.}]$، وهذا يعني إن العينة قد اختيرت من مجتمع فيه $(\rho = \rho_0)$وفق معطيات العينة عند مستوى المعنوية $(\propto)$.

ب- يتم رفض فرضيةالعدم (H_0)،عندما تكون القيمة المطلقة لإحصاءة الاختبار المحسوبة أكبر من أو تساوي القيمة الجدولية، أي إن $[|Z_{cal.}| \geq Z_{tab.}]$، وهذا يعني إن العينة لم يتم اختيارها من مجتمع فيه $(\rho = \rho_0)$وفق معطياتها، عند مستوى المعنوية$(\propto)$.

➢ **مثال :**

إذا كانت قيمة معامل الارتباط البسيط لبيرسون (r_p)، بين متغيري الوزن (X)و الطول (Y)لخمسة طلاب، مساوية الى$(r_p = 0.92)$.

المطلوب :

هل يمكن القول بأن هذه العينة قد اختيرت من مجتمع فيه معامل الارتباط البسيط بين المتغيرين مساوي الى(0.95)عند مستوى المعنوية($\propto = 1\%$)؟

الحل :

للإجابة عن السؤال السابق، نقوم باختيار الفرضية الاحصائية الآتية :

$$H_0: p = 0.95$$
$$H_1: p \neq 0.95$$

نقوم بإيجاد إحصاءة الاختبار (Z).

$$Z_{cal.} = \frac{w - E(w)}{\sqrt{Var(w)}}$$

$$w = \frac{1}{2} Ln\left(\frac{1+r}{1-r}\right)$$

$$= \frac{1}{2} Ln\left(\frac{1+0.92}{1-0.92}\right)$$

$$= \frac{1}{2} Ln(24)$$

$$= 1.589$$

$$E(w) = \frac{1}{2} Ln\left(\frac{1+\rho_0}{1-\rho_0}\right)$$

$$= \frac{1}{2} Ln\left(\frac{1+0.95}{1-0.95}\right)$$

$$= \frac{1}{2} Ln(39)$$

$$= 1.832$$

$$Var(w) = \frac{1}{n-3}$$

$$= \frac{1}{5-3}$$

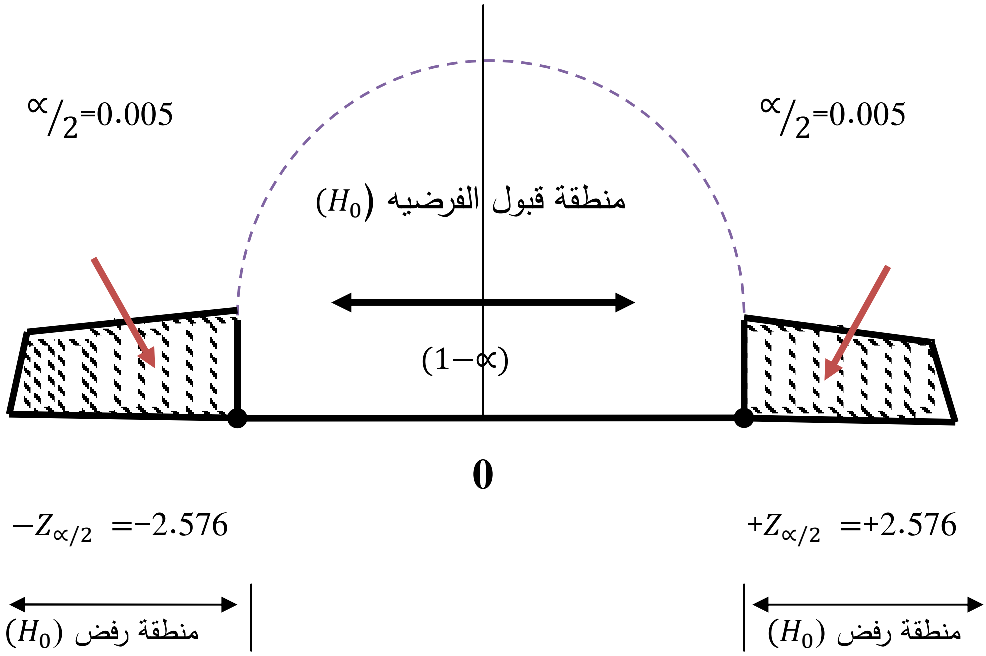
$$= 0.5$$

$$\therefore Z_{cal.} = \frac{1.589 - 1.832}{\sqrt{0.5}}$$

$$= \frac{-2.43}{0.707}$$

$$= -0.344$$

بما أن الاختبار ذو جانبين، تحت مستوى المعنوية $(\propto = 0.01)$ لذا سيتم اختيار $(\propto/_2 = 0.005)$ لكل جانب، وبالتالي فإن القيم الجدولية تكون $\left(Z_{\propto/_2} = 2.576\right)$ و $(-Z_{\propto/_2} = 2.576-)$. و الشكل التالي يوضح مناطق رفض وقبول فرضيةالعدم (H_0) :



القرار الاحصائي :

بما أن القيمة المطلقة لإحصاءة الاختبار $(|Z_{cal.}|)$البالغة (0.344)، هي أقل من القيمة الجدولية $(Z_{\propto/2} = 2.576)$ وهذا يعني قبول العدم (H_0)مما يدل ذلك بأن العينة قد اختيرت من مجتمع فيه معامل الارتباط البسيط بين المتغيرين مساو الى $(\rho = 0.95)$وفق معطيات العينة، عند مستوى المعنوية $(\propto = 1\%)$.

مثال :

إذا كان معامل الارتباط البسيط لبيرسون(r_p)بين متغيري الدخل الشهري(X)والانفاق الأسري (Y)، محسوب على أساس عينة عشوائية من الأسر قوامها(150)أسرة، مساو الى$(r_p = 0.85)$.

المطلوب :

هل يمكن القول بأن هذه العينة اختيرت من مجتمع فيه معامل الارتباط البسيط بين المتغيرين يزيد على (0.81). عند مستوى المعنوية$(\propto = 10\%)$؟

الحل :

للإجابة عن السؤال السابق، نقوم باختيار الفرضية الاحصائية الآتية :

$$H_0: p \leq 0.81$$

$$H_1: p > 0.81$$

نقوم بإيجاد إحصاءة الاختبار (Z)، وفقا للصيغة الآتية :

$$Z_{cal.} = \frac{w - E(w)}{\sqrt{Var(w)}}$$

$$w = \frac{1}{2} Ln\left(\frac{1+r}{1-r}\right)$$

$$= \frac{1}{2} Ln\left(\frac{1+0.85}{1-0.85}\right)$$

$$= 1.256$$

$$E(w) = \frac{1}{2} Ln\left(\frac{1+\rho_0}{1-\rho_0}\right)$$

$$= \frac{1}{2} Ln\left(\frac{1+0.81}{1-0.81}\right)$$

$$= 1.127$$

$$Var(w) = \frac{1}{n-3}$$

$$= \frac{1}{150-3}$$

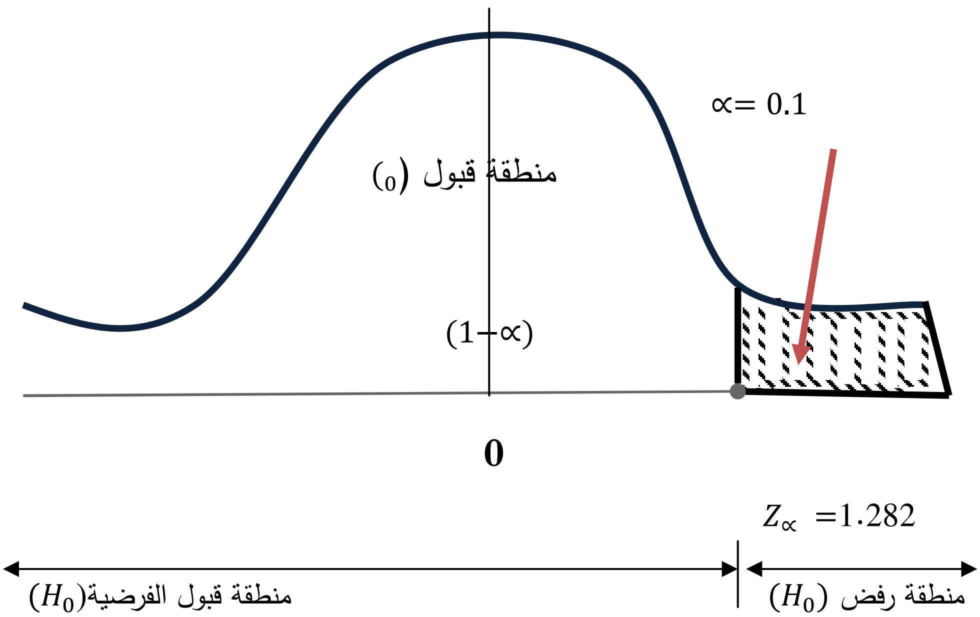
$$= 0.0068$$

$$\therefore Z_{cal.} = \frac{1.256 - 1.127}{\sqrt{0.0068}}$$

$$= \frac{0.129}{0.082}$$

$$= 1.573$$

بما أن الاختبار من جانب واحد، عليه فإن القيمة الجدولية تكون $Z_{\propto} = 1.282$ عند مستوى المعنوية ($\propto = 10\%$) والشكل التالي يوضح مناطق وفض وقبول فرضية العدم (H_0).



القرار الاحصائي :

بما إن القيمة المطلقة لإحصاءة الاختبار ($|Z_{cal.}|$)البالغة (1.573) ، هي أكبر من القيمة الجدولية ($Z_{\propto} = 1.282$)وهذا يعني رفض فرضية العدم($_0$)مما يدل ذلك بأن العينة فد اختيرت من مجتمع فيه معامل الارتباط البسيط بين المتغيرين يزيد على(0.81) بمعنى إن ($\rho > 0.81$)وفق معطيات العينة عند مستوى المعنوية($\propto = 10\%$).

➢ **ثانيا : اختيار معنوية معامل الارتباط عن الصفر ($\rho = 0$) عندما تكون $n < 30$**

لاختيار معنوية معامل الارتباط البسيط لبيرسون(r_p) في هذه الحالة، نقوم باختيار إحدى الفرضيات الاحصائية الثلاث الآنفة الذكر، و منها على وجه التحديد الفرضية الاحصائية الآتية :

$$H_0: \rho = 0$$
$$H_i: \rho \neq 0$$

إن اخصائية الاختبار الملائمة للفرضية أعلاه، تأخذ الشكل الآتي :

$$\mathrm{t}_{cal.} = r^* \sqrt{\frac{n-2}{1= r^2}} - t(n-2)$$

حيث إن :

r: تمثل معامل الارتباط البسيط لبيرسون. $r^* = r$

n : أزواج قيم المتغيرين (X) و (Y).

ولاتخاذ القرار الاحصائي حول رفض أو عدم رفض فرضية العدم (H_0)، يتم مقارنة القيمة المطلقة لاحصاءة الاختيار المحسوبة $(|t_{cal.}|)$ مع القيمة الجدولية (t_{tab})، المتحصل عليها من جداول توزيع (t)، اعتمادا على درجات الحرية $(n-2)$، و مستوى المعنوية $(\propto)$، ونوع الفرضية البديلة (H_1).

القرار الاحصائي :

أ- يتم قبول فرضية العدم (H_0)، عندما تكون القيمة المطلقة لإحصاءة الاختبار المحسوبة أقل من القيمة الجدولية، أي إن $\left[|t_{cal.}| < t_{(n-2)}\right]$، مما يدل على عمد وجود علاقة معنوية بين المتغيرين، وفق معطيات العينة، عند مستوى المعنوية $(\propto)$.

ب-يتم رفض فرضية العدم (H_0)، عندما تكون القيمة المطلقة لاحصاءة الاختبار أكبر من أو تساوي القيمة الجدولية، أي إن $\left[|t_{cal.}| \geq t_{(n-2)}\right]$، مما يدل على وجود علاقة معنوية بين المتغيرين، وفق معطيات العينة، عند مستوى المعنوية $(\propto)$.

مثال :

إذا كان معامل الارتباط البسيط لبيرسون (r_p)، بين مصروف جيب الطالب (X) وعلامته الفصلية (Y) في مساق الرياضيات، محسوب على أساس عينة عشوائية من الطلاب قوامها (8) طلاب مساوي الى $(r_p = 0.97)$.

المطلوب :

هل أن قيمة معامل الارتباط البسيط المحسوبة، تدل على وجود علاقة معنوية بين المتغيرين وفق معطيات العينة، عند مستوى المعنوية $(\propto = 0.05)$؟

الحل :

للإجابة عن السؤال السابق، نقوم باختيار الفرضية الاحصائية الآتية :

$$H_0: \rho = 0$$
$$H_1: \rho \neq 0$$

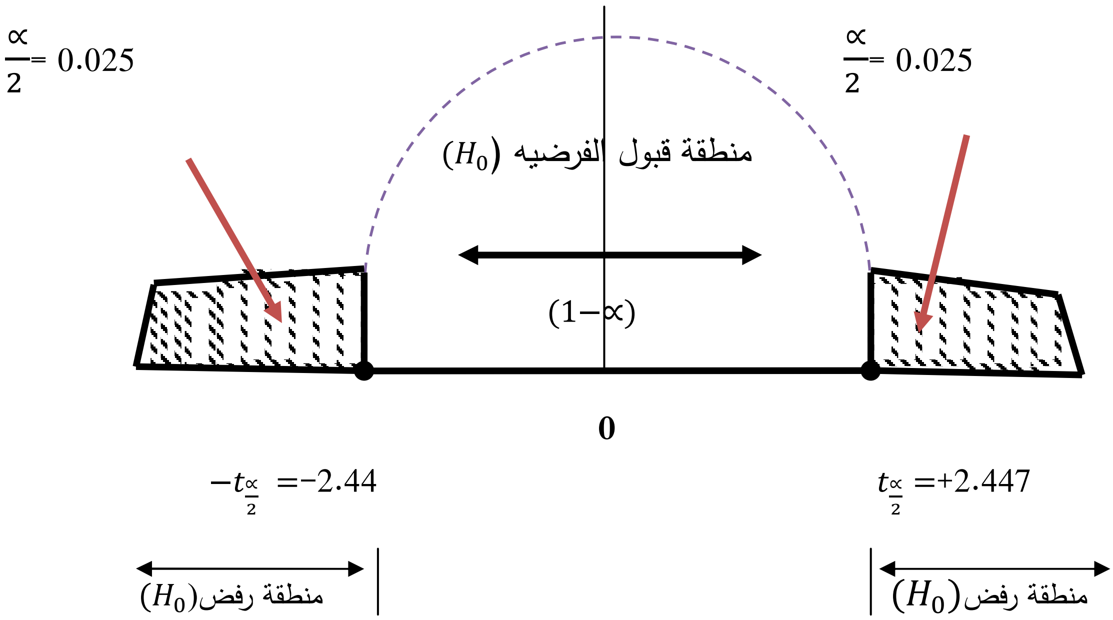
نقوم بحساب إحصاءة الاختبار (t)، وفقا للصيغة الآتية :

$$t_{cal.} = r^{*}\sqrt{\frac{n-2}{1-r^{2}}}$$

$$= (0.97)^{*}\sqrt{\frac{8-2}{1-(-0.97)^{2}}}$$

$$= 9.774$$

بما إن الاختبار ذو جانبين، تحت مستوى المعنوية $(\propto = 0.05)$، لذا سيتم اختبار $\left(\frac{\propto}{2}\right)$ لكل جانب، و بالتالي فإن قيم(t) الجدولية، تكون $\left(t_{\frac{\propto}{2}} = 2.447\right)$ و $\left(-t_{\frac{\propto}{2}} = 2.447\right)$ بدرجة حرية 6 والشكل التالي يوضح مناطق رفض قبول فرضية العدم (H_0):



القرار الاحصائي :

بما إن القيمة المطلقة لاحصاءة الاختبار المحسوبة $(|t_{cal.}|)$ البالغة (9.774)، هي أكبر من القيمة الجدولية البالغة (2.447)، وهذا يعني رفض فرضية العدم (H_0)،مما يدل ذلك على وجود علاقة معنوية بين المتغيرين، أي إن $(\rho \neq 0)$ وفق معطيات العينة عند مستوى المعنوية $(\propto = 0.05)$.

4- الارتباط الجزئي :Partial Correlation

يلاحظ أحيانا عن وجود ارتباط بين متغيرين يعزي جزئيا الى ارتباط متغير ثالث مرتبط مع كليهما مثال ذلك : العلاقة ما بين دخل الأسرة الشهري(X_1) وإنفاقها الشهري (X_2) يتأثر بعلاقة عدد أفراد الأسرة (X_3) مع كل من المتغيرين (X_1) و (X_2).

ولقياس درجة الارتباط بين أي متغيرين باستبعاد أثر المتغير الثالث نستخدم لذلك ما يسمى بمعامل الارتباط الجزئي (**Partial CorrelationCoefficient**)عليه فإن معامل الارتباط الجزئي بين المتغيرين ((X_1 و X_2) باستبعاد أثر المتغير (X_3) يكتب على الوجه الآتي :

$$r_{12.3} = \frac{r^{12} - r^{13} - r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}}$$

مثال :

إذا كان لديك مصفوفة معاملات الارتباط البسيط الآتية :

$$\mathrm{R} = \begin{bmatrix} 1 & 0.91 & 0.39 \\ & 1 & 0.62 \\ & & 1 \end{bmatrix}$$

حيث إن :

X_1 : دخل الأسرة الشهري.

X_2 : إنفاق الأسرة الشهري.

X_3 : عدد افراد الأسرة.

المطلوب :

1- إيجاد معامل الارتباط الجزئي بين دخل الأسرة (X_1) وإنفاقها الشهري (X_2) باستبعاد أثر عدد أفراد الأسرة (X_3).

2- حساب معامل الارتباط الجزئي بين إنفاق الأسرة (X_2) وعدد أفرادها (X_3) باستبعاد أثر دخل الأسرة (X_1).

الحل :

-1 إيجاد معامل الارتباط الجزئي $[r_{23.1}]$:

$$r_{12.3}=\frac{r_{12}-r_{13}-r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}}$$

$$=\frac{0.91-0.39(0.62)}{\sqrt{1-(0.39)^2}\sqrt{1-(0.62)^2}}$$

$$=0.93$$

من أعلاه يتضح بأن العلاقة طردية وقوية بين دخل الأسرة (X_2) و إنفاقها الشهري (X_3) بعد استبعاد أثر عدد أفراد الأسرة (X_3).

-2 ايجاد معامل الارتباط الجزئي $[r_{23.1}]$:

$$r_{23.1}=\frac{r_{23}-r_{12}-r_{13}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{13}^2}}$$

$$=\frac{0.62-0.91(0.39)}{\sqrt{1-(0.91)^2}\sqrt{1-(0.39)^2}}$$

$$=0.69$$

يتضح من النتيجة السابقة، بأن العلاقة طردية و متوسطو بين إنفاق الأسرة (X_2) وعدد أفرادها(X_3) بعد إستبعاد أثر دخل الأسرة (X_1).

-5 اختبار معنوية معامل الارتباط الجزئي :

على إفتراض لدينا $(X_1,X_2,\dots.X_k)$ تمثل متغيرات عشوائية عددها(k) ذات توزيع طبيعي متعدد المتغيرات (Multivariate Normal Distribution).

وإن **يمثل معامل الارتباط الجزئي بين المتغيرين** (X_i) $(\boldsymbol{r_{ij.}\,123\,\dots.p})$ و (X_j)بعد استبعاد أثر (p) من المتغيرات الأخرى، حيث إن $(\rho\le k-2)$.

لاختبار معنوية معامل الارتباط الجزئي $(\boldsymbol{r_{ij.}}\,\dots..)$دعنا نضع فرضيتا الاختبار الآتيتين :

$$H_0: \rho_{ij} = 0$$
$$H_1: \rho_{ij} \neq 0$$

إن احصاءة اختبار الفرضية أعلاه هي :

$$t = (r_{ij........}) \sqrt{\frac{n-p-2}{1-r_{ij.}^2}} \sim t_{(n-p-2)}$$

حيث إن :

$r_{ij.}$: معامل الارتباط الجزئي بين المتغيرين (X_i) و (X_j) بعد استبعاد أثر المتغيرات الأخرى.

p : عدد المتغيرات التي تم أستبعاد أثرها.

n : عدد مفردات العينة المختارة.

القرار الاحصائي :

أ- يتم قبول فرضية العدم (H_0)، عندما تكون القيمة المطلقة لإحصاءة الاختبار المحسوبة أقل من القيمة الجدولية، أي إن $[|t_{cal.}| < t_{tab.}]$، وهذا يعني إن قيمة معامل الارتباط الجزئي $[r_{ij.....}]$غير معنوية وفق معطيات العينة، عند مستوى المعنوية $(\propto)$.

ب-يتم رفض فرضية العدم (H_0)، عندما تكون القيمة المطلقة لاحصاءة الاختبار المحسوبة أكبر من أو تساوي القيمة الجدولية، أي إن $[|t_{cal.}| > t_{tab.}]$، وهذا يعني إن قيمة معامل الارتباط الجزئي $[r_{ij.....}]$تكون معنوية وفق معطيات العينة، عند مستوى المعنوية $(\propto)$.

مثال :

إذا توفرت لديك المعلومات الآتية :

$$r_{12.3} = 0.7 \qquad , n = 20$$

المطلوب :

اختبر الفرضية الاحصائية التالية، مستخدما مستوى المعنوية$(\propto = 0.05)$.

$$H_0: \rho_{12.3} = 0$$

$$H_1: \rho_{12.3} \neq 0$$

الحل :

لاختبار الفرضية الاحصائية السابقة، نقوم بإيجاد إحصاءة الاختبار (t)، وفقا للصيغة الآتية :

$$t_{cal.} = (r_{12.3}) \frac{n - \rho - 2}{\sqrt{1 - r_{12.3}^2}}$$

$$= (0.7) \sqrt{\frac{20 - 1 - 2}{1 - (0.7)^2}}$$

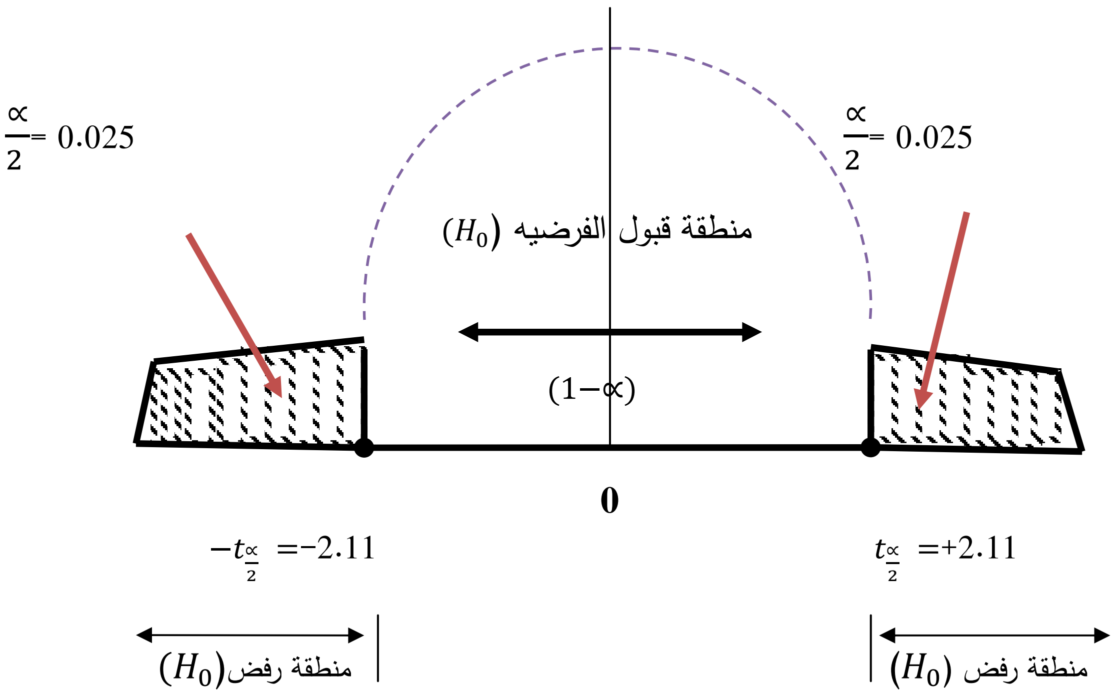
$$= (0.7) \sqrt{\frac{17}{1 - 0.49}}$$

$$= (0.7) \sqrt{\frac{17}{0.51}}$$

$$= (0.7) \sqrt{33.33}$$

$$=4.04$$

بما إن الاختبار ذو جانبين، تحت مستوى المعنوية $(\propto = 0.05)$، لذا سيتم اختبار $\left(\frac{\propto}{2}\right)$لكل جانب، عليه تكون قيم (t) الجدولية بدرجة حرية (17) عند مستوى المعنوية المذكور كالآتي $[t_{\propto/2} = 2.11]$و $[-t_{\propto/2} = 2.11]$، و الشكل التالي يوضح مناطق رفض وقبول فرضية العدم (H_0):



القرار الاحصائي :

بما أن القيمة المطلقة لإحصاءة الاختبار المحسوبة ($|t_{cal.}|$) البالغة (4.04)، هي أكبر من القيمة الجدولية البالغة (2.11)، عليه سيتم رفض فرضية العدم (H_0) وهذا يعني إن قيمة معامل الارتباط الجزئي [$r_{12.3}$] تكون معنوية، عند مستوى المعنوية ($\propto = 0.05$).

6- الارتباط المتعدد: MultipleCorrelation

في أغلب الأحيان نجد بأن التغير الذي يطرأ على ظاهرة ما ناجم عن تغير مجموعة ظواهر أخرى مجتمعة وليس ظاهرة واحدة فقط.

فعلى سبيل المثال : إن إنتاجية القمح تتأثربنوع البذور المستخدمة و نوع السماد المستخدم، و نوع التربة .

ولقياس درجة العلاقة بين متغير معين ومجموعة متغيرات أخرى نستخدم لذلك معامل الارتباط المتعدد **(MultipleCorrelation Coefficient)** فعلى سبيل المثال يمكن قياس درجة العلاقة بين المتغير (X_1, X_3, X_2) وفقا للصيغة الآتية :

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2(r_{12})(r_{13})(r_{23})}{1 - r_{23}^2}}$$

مثال :

إذا كان لديك مصفوفة معاملات الارتباط البسيط الآتية :

$$R = \begin{bmatrix} 1 & 0.91 & 0.39 \\ & 1 & 0.62 \\ & & 1 \end{bmatrix}$$

حيث إن :

X_1 : إنفاق الاسرة.

X_2 : دخل الأسرة.

X_3 : عدد افراد الأسرة.

المطلوب :

إيجاد معامل الارتباط التعدد بين إنفاق الأسرة(X_1) وكل من الدخل(X_2)و عدد أفراد الأسرة (X_3).

الحل :

إيجاد معامل الارتباط $R_{1.23}$ نقوم بتطبيق الصيغة الآتية :

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2(r_{12})(r_{13})(r_{23})}{1 - r_{23}^2}}$$

$$= \frac{(0.91)^2 + (0.39)^2 - 2(0.91)(0.39)(0.62)}{1 - (0.62)^2}$$

$$=0.94$$

من النتيجة السابقة يتضح، بأن العلاقة طردية و قوية جداً بين إنفاق الأسرة (X_1)وكل من دخلها الشهري (X_2) وعدد الأفراد(X_3).

7- اختبار معنوية معامل الارتباط المتعدد :

بافتراض أن $(X_1, X_2, \ldots . X_3)$ تمثل متغيرات عشوائية عددها(k) ذات توزيع طبيعي متعدد المتغيرات (MND).

وإن$(R_{i.123} \ldots \ldots \ldots k)$يمثل تعامل الارتباط المتعدد بين المتغير (X_i) مع بقية المتغيرات الأخرى البالغ عددها $(k-1)$.

لاختبار معنوية معامل الارتباط المتعدد بين المتغير مع بقية المتغيرات الأخرى $(R_{i.})$ دعنا نضع فرضيتا الاختبار الآتيتين :

$$H_0: R_{i.} = 0$$

$$H_1: R_{i.} > 0$$

إن إخصاءة اختبار الفرضية (H_0) أعلاه، هي :

$$F = \frac{R_{i.}^2}{1 - R_{i.}^2} \cdot \frac{n-k}{k-1} \sim F_{[(k-1),(n-k)]}$$

حيث إن :

$r_{i.}$: يمثل معامل الارتباط المتعدد بين المتغير (X_i) مع بقية المتغيرات الأخرى.

k : يمثل عدد المتغيراتا.

n : يمثل عدد مفردات العينة المختارة.

القرار الاحصائي :

أ- يتم قبول فرضية العدم (H_0)، عندما تكون القيمة المطلقة لإحصاءة الاختبار (F) المحسوبة أقل من القيمة الجدولية، أي إن $[F_{cal.} < F_{tab.}]$، وهذا يعني إن قيمة معامل الارتباط المتعدد $(R_{i.})$ غير معنوية وفق معطيات العينة، عند مستوى المعنوية $(\propto)$.

ب-يتم رفض فرضية العدم (H_0)، عندما تكون القيمة المطلقة لاحصاءة الاختبار(F)المحسوبة أكبر من أو تساوي القيمة الجدولية، أي إن $[F_{cal.} \geq F_{tab.}]$، وهذا يعني إن قيمة معامل الارتباط المتعدد $(R_{i.})$تكون معنوية وفق معطيات العينة، عند مستوى المعنوية $(\propto)$.

مثال :

إذا توفرت لديك المعلومات الآتية :

$$R_{1.23} = 0.8 \qquad ,n = 20$$

المطلوب :

اختبر الفرضية الاحصائية التالية، مستخدما مستوى المعنوية$(\propto = 0.05)$.

$$H_0: R_{1.23} = 0$$

$$H_1: R_{1.23} > 0$$

الحل :

لاختبار الفرضية السابقة، نقوم بإيجاد إحصاءة الاختبار(F)، كالآتي :

$$F = \frac{R_{1.23}^{2}}{1 - R_{1.23}^{2}}, \frac{n-k}{k-1}$$

$$= \frac{(0.8)^2}{1-(08)^2}, \frac{20-3}{2}$$

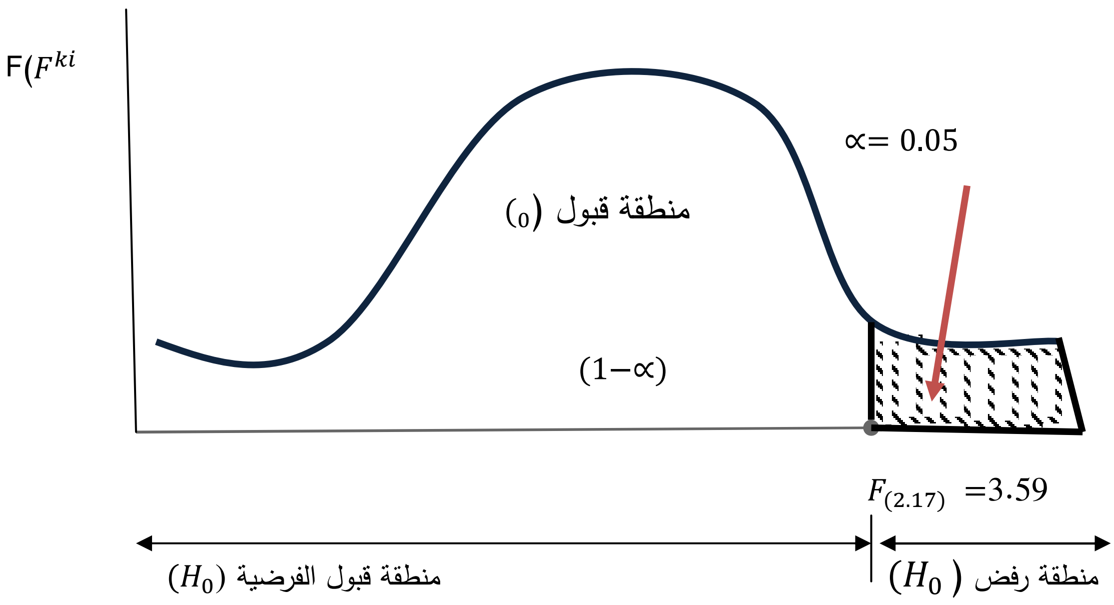
$$= \frac{0.64}{1-0.64}, \frac{17}{2}$$

$$= \frac{17(0.64)}{2(0.36)}$$

$$= \frac{10.88}{0.72}$$

$$=15.11$$

من جداول توزيع(F) بدرجتي حرية (17,2)، وعند مستوى معنوية ($\propto= 0.05$)، نجد إن قيمة ($F_{\propto}$) الجدولية هي (3.59)، و الشكل التالي يوضح مناطق وفض وقبول فرضية العدم (H_0)



القرار الاحصائي :

بما إن قيمة إحصاءة الاختبار(F)المحسوبة البالغة (15.11)، هي أكبر من قيمة ($F_{\propto}$) الجدولية البالغة (3.59)، عليه سيتم رفض فرضية العدم (H_0)، وهذا يعني إن قيمة معامل الارتباط المتعدد [$R_{1.23}$] تكون معنوية، عند مستوى المعنوية ($\propto= 0.05$).

8- ارتباط الرتب: Ranks Correlation

في العناصر السابقة ثم توضيح مفهوم الارتباط الخطي البسيط والارتباط الجزئي و المتعدد، وأهم الصيغ المستخدمة في حساب معاملات الارتباط المذكورة بهدف قياس العلاقة الارتباطية بين متغيرين أو أكثر من النوع الكمي (أي يمكن قياسهما كمياً)، إلا إنه في بعض الأحيان يكون المتغيرين من النوع الوصفي (أي لايمكن قياسهما كمياً) من الناحية العملية، مما يتعذر علينا استخدام صيغ معامل الارتباط البسيط لبيرسون لقياس العلاقة بين المتغيرين الوصفيين، لذا نلجأ إلىإجراء بعض التحويلات على المتغيرين الوصفيين، وتتلخص هذه التحويلات باستخدام الرتب **(Ranks)** بدلاً من القيم الأصلية توضع مقابل قيم المتغيرين الوصفيين بعد أن يتم ترتيب قراءات كل منهما تصاعدياً أو تنازلياً كل على إنفراد، مع مراعاة أخذ متوسط الرتب في حالة وجود قراءات متكررة تحمل نفس الصفة للمتغير الوصفي.

وتكون معاملات ارتباط الرتب على نوعين مهمين، هما :

أ- معاملارتباط الرتب لسبيرمان**Spearman's Rank Correlation Coefficient**

ب-معامل ارتباط الرتب لكندال.**Kendall's Rank correlation coefficient**

وسيتم التركيز بشكل منفصل في هذاالمطبوع على دراسة معامل ارتباط الرتب لسبيرمان، لأهميته في قياس العلاقة الارتباطية بين متغيرين من النوع الوصفي، مما يجعل هذا المعامل أكثر شيوعاً واستخداماً من معامل ارتباط الرتب لكندال.

معامل ارتباط الرتب لسبيرمان :

وهو مؤشر احصائي يستخدم لقياس العلاقة الارتباطية بين متغيرين وصفيين (أي لا يمكن قياسهما كمياً) مثال ذلك : قياس العلاقة بين مستوى ذكاء الطالب (X) ومستوى أدائه العلمي (Y) أو قياس العلاقة بين متغيرين أحدهما وصفي والأخر كمي، كما يمكن استخدامه لقياس العلاقة بين متغيرين كميين، إلا إنه يفضل في هذه الخالة استخدام معامل الارتباط لبيرسون، لدقته كونه يتعامل مع القيم الأصلية وليس الرتب عند ايجاد معامل ارتباط الرتب لسبيرمان**(Spearman)** يتم أولاً إعطاء رتب **(Ranks)** مقابلة لقيم المتغيرات الوصفية أو الكمية بعد أن يتم ترتيب القراءات تصاعدياً أو تنازلياً لكل متغير كل على أنفراد، مع مراعاة أخذ متوسط الرتب في حالة وجود قراءات تحمل نفس الصفة للمتغير الوصفي أو تأخذ نفس القيمة بالنسبة للمتغير الكمي.

ويٌعد معامل ارتباط الرتب لسبيرمان من أكثر معاملات ارتباط الرتب شيوعاً واستخداماً و الصيغة العامة لمعامل ارتباط الرتب لسبيرمان، تأخذ الشكل الآتي :

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

حيث إن :

n : يمثل عدد أزواج القيم(Y_i , X_i) لكلا المتغيرين.

d_i : يمثل الفرق بين رتب المتغير (X_i) ورتب المتغير(Y_i)، أي إن $[d_i = R_x - R_y]$.

مثال :

البيانات التالية، تمثل معدلات (8) ثمانية طلاب في الثانوي(X_i)وتقديراتهم في الجامعة (Y_i)

المعدل في الثانوي (X_i)	59	72	61	80	86	82	90	67
التقدير في الجامعة (Y_i)	مقبول	جيد	متوسط	جيد	جيد جداً	جيد	امتياز	مقبول

المطلوب :

قياس العلاقة الارتباطية بين معدل الطالب في الثانوي(X_i)، و تقديره في الجامعة (Y_i) باستخدام معامل ارتباط الرتب لسبيرمان.

الحل :

نقوم بحساب المعلومات التالية، و الموضحة بالجدول الآتي :

X_i	Y_i	$Rank(X_i)$	$Rank(Y_i)$	$d_i = R_x - R_y$	d_i^2
59	مقبول	1	1.5	-0.5	0.25
72	جيد	4	5	-1	1
61	متوسط	2	3	-1	1
80	جيد	5	5	0	0
86	جيد جداً	7	7	0	0
82	جيد	6	5	+1	1
90	امتياز	8	8	0	0
67	مقبول	3	1.5	-1.5	2.25
-	-	-	-	0	5.5
					$\sum d_i^2$

$$\therefore r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

$$= 1 - \frac{6(5.5)}{8(64-1)}$$

$$= 1 - \frac{33}{504}$$

$$= 1 - 0.07$$

$$= +0.93$$

من النتيجة السابقة، يتضح بأن العلاقة بين معدل الطالب في الثانوي(X_i)، وتقديره في الجامعة (Y_i)، هي علاقة طردية وقوية جداً، وهذا يعني بأن الطالب عندما يكون متفوقاً في مرحلة الثانوي سيقود الى تفوقه في الدراسة الجامعية.

خصائص معامل ارتباط الرتب لسبيرمان :

يتصف معامل ارتباط الرتب لسبيرمان، بالخصائص الآتية :

1. يستخدم معامل ارتباط الرتب لسبيرمان، لقياس العلاقة الارتباطية بين المتغيرات الوصفية (التي لا يمكن قياسها كمياً) ويمكن استخدامه أحياناً لقياس العلاقة بين المتغيرات الكمية، بعد التعبير عن القيم الأصلية للمتغيرين بدلالة الرتب.
2. يٌعد معامل ارتباط الرتب لسبيرمان، سهل الحساب و الفهم والتطبيق مقارنة بمعامل الارتباط السيط لبيرسون.
3. إن قيمة معامل ارتباط الرتب لسبيرمان، تقع ضمن المجال $(-1 \le r_s \le +1)$، إذ إن :
   أ- عندما يكون $(\sum d_i^2 = 0)$ $\Longleftarrow$ فإن $(r_s = +1)$.
   ب- عندما يكون $[6\sum d_i^2 = 2n(n^2-1)]$ $\Longleftarrow$ فإن $(r_s = -1)$.
   ت- عندما يكون$[6\sum d_i^2 = 2n(n^2-1)]$ $\Longleftarrow$ فإن $(r_s = 0)$.
4. إن قيمة معامل ارتباط الرتب لسبيرمان المحسوبة في حالة المتغيرات الكمية لا تساوي بالضبط قيمة معامل الارتباط البسيط لبيرسون، وذلك بسبب التعامل مع رتب القيم وليس مع القيم الأصلية للمتغيرين.
5. إن معامل ارتباط الرتب لسبيرمان مشتق من معامل الارتباط البسيط لبيرسون، بعد التعبير عن القيم الأصلية للمتغيرين بدلالة الرتب ***(Ranks)***.

اختبار معنوية معامل ارتباط الرتب لسبيرمان :

بافتراض إن (r_s) يمثل معامل ارتباط الرتب لسبيرمان بين أزواج قيم متغيرين وصفيين، أو أحدهما كمي و الأخر وصفي، محسوب على اساس عينة عشوائية قوامها (n)، فإن اختبار معنوية معامل ارتباط الرتب لسبيرمان يكون على ثلاثة أشكال و على النحو الآتي :

- **أولاً : الاختبار الرتبي اللامعلمي :**

يستخدم الاختبار الرتبي اللامعلمي، لاختبار معنوية معامل ارتباط الرتب لسبيرمان (r_s)، عندما يكون عدد أزواج قيم المتغيرين (X)و(Y)، أقل من أو يساوي (10) أزواج، أي إن $(n \leq 10)$ وسيتم شرح هذا الاختبار بشيء من التفصيل في الفصل السابع الخاص بالاختبارات اللامعلمية.

- **ثانياً: اختبار (t):**

يمكن اختبار معنوية معامل ارتباط الرتب لسبيرمان(r_s)، باستخدام واختبار (t)، عندما يكون عدد أزواج قيم المتغيرين (X)و(Y)، تتراوح بين $(10 < n \leq 30)$، ويتم ذلك من خلال اختبار الفرضية الاحصائية الآتية :

$$H0: \rho_s = 0$$
$$H1: \rho_s \neq 0$$

إن إحصاءة الاختبار الملائمة للفرضية أعلاه، تعطي بالصيغة الآتية :

$$t_{cal.} = r^* \sqrt{\frac{n-2}{1-r^2}} \sim t_{(n-2)}$$

مثال :

إذا كان معامل ارتباط الرتب لسبيرمان(r_s)، بين معدل الطالب في الثانوي(X) وتقديره في الجامعة للسنة الأولى (Y)، محسوبة على أساس عينة عشوائية قوامها (20) طالب، مساو الى $(r_s = 0.95)$.

المطلوب :

اختبر معنوية معامل ارتباط الرتب لسبيرمان(r_s)، عند مستوى المعنوية $(\propto = 1\%)$.

الحل :

اختبر معنوية معامل ارتباط الرتب لسبيرمان(r_s)، نقوم باختبار الفرضية الاحصائية الآتية :

$$H0: \rho_s = 0$$
$$H1: \rho_s \neq 0$$

ولاختبار الفرضية السابقة، نقوم بحساب إحصاءة الاختبار (t)، وفقاً للصيغة الآتية :

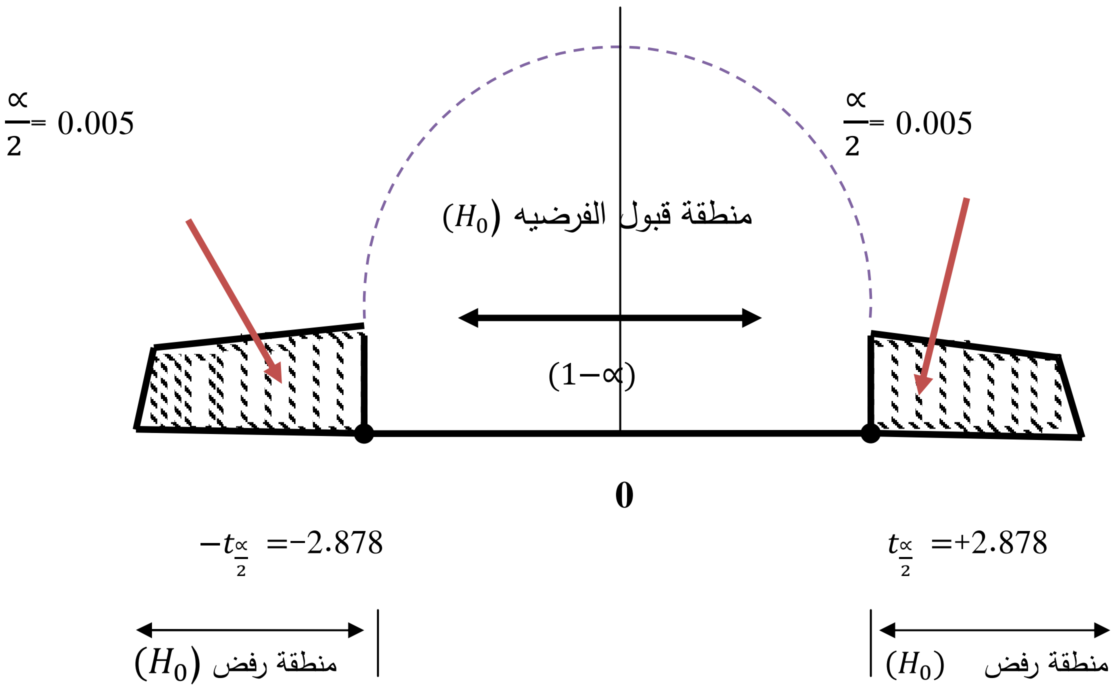
$$t_{cal.} = r^* \sqrt{\frac{n-2}{1-r^2}}$$

$$= 0.95^* \sqrt{\frac{20-2}{1-(0.95)^2}}$$

$$= 0.95^*(13.587)$$

$$=12.908$$

من جداول توزيع (t)، بدرجة حرية (18)، ومستوى المعنوية (0.01)، فإن قيم (t)، الجدولية هي $\left(t_{\propto/2} = 2.878\right)$ و $\left(-t_{\propto/2} = 2.878\right)$، و الشكل التالي يوضح مناطق رفض وقبول فرضية العدم (H_0).



القرار الاحصائي :

بما أن القيمة المطلقة لإحصاءة الاختبار المحسوبة$|t_{cal.}|$)البالغة (12.908) هي أكبر من القيمة الجدولية $\left(t_{\frac{\propto}{2}} = 2.878\right)$، وهذا يعني رفض فرضية العدم $_0$ مما يدل على وجود علاقة معنوية بين المتغيرين، أي إن$(\rho_s \neq 0)$ وفق معطيات العينة، عند مستوى المعنوية $(\propto = 1\%)$.

➢ **ثالثاً : اختبار (Z) :**

يستخدم اختبار (Z)، لاختبار معنوية معامل ارتباط الرتب لسبيرمان(r_s)، عندما يكون عدد أزواج قيم المتغيرين (X)و(Y)، أكثر من (25) زوج، أي إن $(n > 25)$وفي هذه الحالة يكون توزيع معامل ارتباط الرتب لسبيرمان(r_s)، يقترب من التوزيع الطبيعي بمتوسط $(\rho_s = 0)$ وتباين مساو الى$\left(\frac{1}{n-1}\right)$أي إن $[r\_s \sim N(0,1/(n-1))]$.

ولاختبار معنوية معامل ارتباط الرتب السبيرمان، نقوم باختبار الفرضية الاحصائية الآتية :

$$H_0: \rho_s = 0$$
$$H_1: \rho_s \neq 0$$

إن إحصاءة الاختبار الملائمة للفرضية السابقة، تأخذ الشكل الآتي :

$$Z_{cal.} = \frac{r_s - E(r_s)}{\sqrt{Var(r_s)}}$$

$$= r_s{}^* \sqrt{n-1}$$

المثال :

إذا كان معامل ارتباط الرتب لسبيرمان(r_s)، بين متغيري المؤهل العلمي (X)ومستوى الأداء(Y) في إحدى المؤسسات التربوية، محسوب على اساس عينة عشوائية قوامها (50) منتسب مساو الى$(r_s = 0.9)$.

المطلوب :

اختبر معنوية معامل ارتباط الرتب لسبيرمان(r_s)، عند مستوى المعنوية $(\propto = 5\%)$.

الحل :

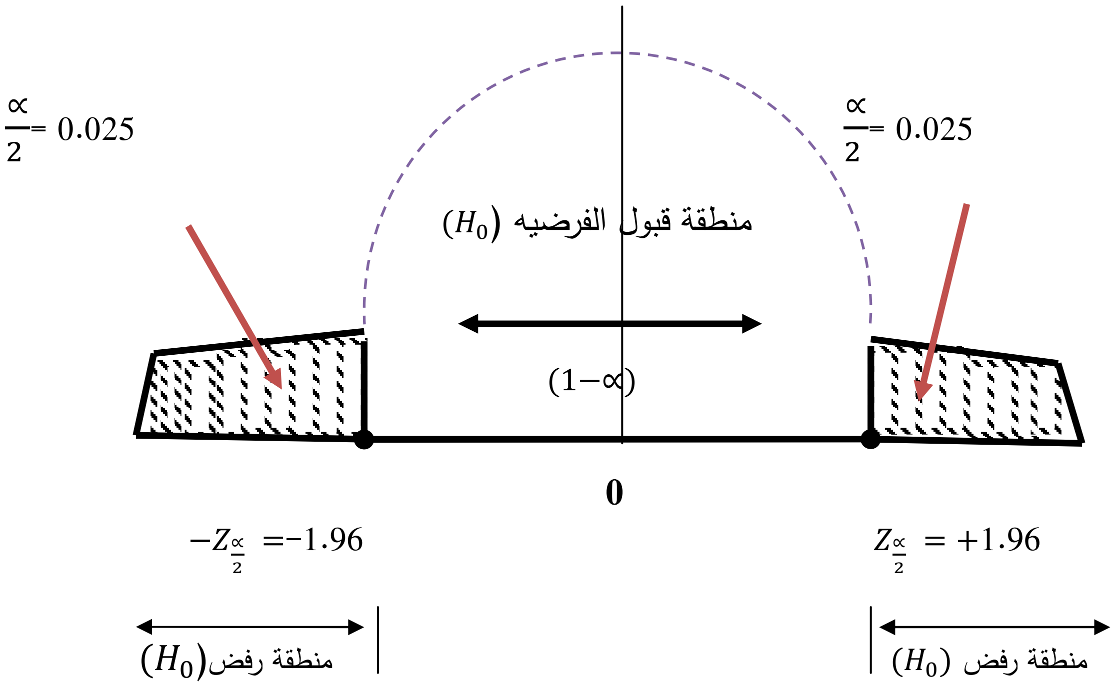
لاختبار معنوية معامل ارتباط الرتب لسبيرمان(r_s)، نقوم باختبار الفرضية الاحصائية الآتية :

$$H_0: \rho_s = 0$$
$$H_1: \rho_s \neq 0$$

ولاختبار الفرضية السابقة نقوم بحساب إحصاءة الاختبار (Z)، وفقاً للصيغة الآتية :

$$Z_{cal.} = r_s{}^* \sqrt{n-1}$$
$$= 0.9 \ {}^* \sqrt{50-1}$$
$$= 0.9^*(7)$$
$$= 6.3$$

بما أن الاختبار ذو جانبين، تحت مستوى المعنوية (0.05)، عليه فإن القيم الجدولية تكون $\left(Z_{\frac{\propto}{2}} = 1.96\right)$ و ، $\left(-Z_{\frac{\propto}{2}} = -1.96\right)$، و الشكل التالي يوضح مناطق رفض وقبول فرضية العدم (H_0).



القرار الاحصائي :

بما أن القيمة المطلقة لإحصاءة الاختبار المحسوبة$|Z_{cal.}|$)البالغة (6.3) هي أكبر من القيمة الجدولية $\left(Z_{\frac{\propto}{2}} = 1.96\right)$، وهذا يعني رفض فرضية العدم (H_0) مما يدل ذلك على وجود علاقة معنوية بين المتغيرين، أي إن $(\rho_s \neq 0)$ وفق معطيات العينة، عند مستوى المعنوية $(\propto = 5\%)$.

تحليل الانحدار Analysis Of Regression

مقدمة

يٌعد الانحدار من المواضيع الأساسية وجزءاً مهماً من النظرية الاحصائية و يتميز الانحدار باستخداماته الواسعة في مختلف العلوم الطبيعية و الإدارية والاقتصادية، فعلى سبيل المثال لا الحصر في المجال الاقتصادي، يٌعد الانحدار الأداةالعلمية التحليلية في الاقتصاد الكلي التحليلي، و القياس الاقتصادي، إذ يمكن استخدامه للتعبير عن العلاقات التي تربط المتغيرات الاقتصادية فيما بينها بصيغة نماذج رياضية يطلق عليها بـ (نماذج الانحدار)،ومن تم تقدير معلمات هذه النماذج واعتمادها

لأغراض عملية التنبؤ بأحد المتغيرات باعتباره متغيراً تابعاً **(Dependent Variable)** عند مستويات محددة لمتغيرات أخرى يطلق عليها بالمتغيرات المستقلة **(Independent Variable).**

وبصورة عامة، يعرف الانحدار بأنه : "أسلوب رياضي لتقدير العلاقة بين متغيرين أو أكثر بدلالة وحدات قياس المتغيرات المعتمدة (التابعة) في العلاقة، وغالباً ما تسمى هذه العلاقات بنماذج الانحدار **(RegressionModels).**

يمكن تقسيم الانحدار من حيث التحليل **(Analysis)** الى قسمين مهمين، هما :

أ- الانحدار الخطي. **LinearRegression**

ب-الانحدار غير الخطي **Non-LinearRegression**

وسيتم التركيز في هذا المطبوع على النوع الأول المتمثل بالانحدار الخطي بشيء من التفصيل لأهميته في تقدير العلاقة الخطية بين متغيرين أو أكثر، و على النحو الآتي :

1- الانحدار الخطي:LinearRegression

إن الهدف الرئيسي من تحليل الانحدار الخطي هو تقدير العلاقة الرياضية الخطية التي تربط بين متغيرين أو أكثر.

الفصل الرابع: تحليل الانحدار

وفي دراسة تحليل الانحدار، يوجد نوعين من المتغيرات، هما :

أ- **المتغير التابع: Dependent Variable**

وهو المتغير الذي تتأثر قيمته في حالة تغير قيمة المتغير المستقل أو (المتغيرات المستقلة)، ولا يتأثر أو تتأثر به.

ب-المتغير المستقل (المتغيرات المستقلة): Independent Variable (s)

وهو المتغير الذي يؤثر في قيمة المتغير التابع عند تغيره، ولا يتأثر بالمتغير التابع، ويسمى أحياناً بالمتغير التفسيري **(Explanatory Variable).**

ويكون الانحدار الخطي على نوعين أساسيين، اعتماداً على عدد المتغيرات الداخلة في النموذج الرياضي للانحدار، هما :

- ✓ **الانحدار الخطي البسيط. Simple LinearRegression**
- ✓ **الانحدار الخطي المتعدد. MultipleLinearRegression**

وفيما يلي شرحاً مفصلاً لكل نوع من أنواع الانحدار الخطي، وعلى النحو الآتي :

2- الانحدار الخطي البسيط: SimpleLinearRegression

يعرف الانحدار الخطي البسيط، بأنه :"عملية تقدير العلاقة الخطية بين متغيرين فقط، أحدهما متغيراً مستقيلاً، والآخر متغيراً (تابعاً).

وعلى فرض إن نموذج الانحدار الخطي البسيط، يأخذ الشكل الآتي :

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad , i = 1,2,3, \dots \dots , n \qquad \dots \dots . (1)$$

حيث إن :

Y_i : المتغير التابع.**(Dependent Variable)**

X_i : المتغير المستقل.**(Independent Variable)**

β_0, β_1: معلمات النموذج.**(Parameters of Model)**

β_0: معامل التقاطع**(Intercept Coefficient).**

β_1 : معامل الانحدار. **(Regression Coefficient).**

ε_i : حد الخطأ العشوائي **(RandomErrorterm).**

ويتصف حد الخطأ العشوائي، بالافتراضات الآتية :

$$(a)E(\varepsilon_i) = 0$$

$$(b)Var\ (\varepsilon_i) = \sigma^2$$

$$(c)Cov(\varepsilon_i, \varepsilon_j) = 0 \quad ,(\forall\ i \neq j).$$

تقدير معلمات نموذج الانحدار الخطي البسيط $(\hat{\beta}_1\ , \beta_0\ )$

إن الهدف من تحليل الانحدار الخطي البسيط، هو تقدير قيم عددية لمعلمات نموذج الانحدار الخطي البسيط $(\hat{\beta}_1\ , \beta_0\ )$

ولتحقيق هذا الغرض نقوم بتطبيق طريقة المربعات الصغرى **(Least Squares)**التي تجعل مجموع مربعات الاخطاء $\left(\sum_i^n \varepsilon_i^2\right)$ اقل ما يمكن **(Minimum)** ويمكن الحصول على الأخطاء أو البواقي من خلال طرح القيم التقديرية $(\hat{Y}_i)$ من القيم الفعلية (Y_i)، أي إن :

$$e_i = Y_i - \hat{Y}_i \qquad \ldots\ldots\ldots\ldots..(\mathbf{2})$$

من المعادلة رقم (2)، يمكن الحصول على مجموع مربعات الأخطاء، كالآتي :

$$\sum_i^n e_i^2 = \sum_i^n (Y_i - \hat{Y}_i)^2 \qquad \ldots\ldots\ldots\ldots.(\mathbf{3})$$

إن تطبيق طريقة المربعات الصغرى (LS) على المعادلة رقم (3) تنطوي على تحديد قيمة كل من $(\hat{\beta}_1\ , \beta_0\ )$ التي تجعل مجموع مربعات الأخطاء أقل ما يمكن $\left(\sum_i^n e_i^2\right)$ (Minimum)

وعلى النحو الآتي :

$$\sum_i^n e_i^2 = \sum_i^n (Y_i - \hat{\beta}_0 - \hat{\beta}_i X_i)^2 \Rightarrow Minimize \ldots\ldots\ldots..(\mathbf{4})$$

وبإجراء التفاضل الجزئي على المعادلة رقم (4) بالنسبة للمعلمتين $(\hat{\beta}_1 ,\beta_0\ )$ نحصل على :

$$\frac{\sum_i^n e_i^2}{\partial\hat{\beta}_0} = -2\sum_i^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \qquad \ldots\ldots\ldots\ldots(\mathbf{5})$$

$$\frac{\sum_{i}^{n} e_i^2}{\partial\hat{\beta}_1} = -2x_i \sum_{i}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0 \qquad \ldots\ldots\ldots\ldots (6)$$

وبتبسيط العلاقتين رقم (5) و (6)، نحصل على المعادلتين الطبيعيتين**(Tow Normal Equations)** على النحو الآتي :

$$\sum_{i}^{n} Y_i = n\,\hat{\beta}_0 + \hat{\beta}_1 \sum_{i}^{n} X_i \qquad \ldots\ldots\ldots\ldots (7)$$

$$\sum_{i}^{n} X_i Y_i = \hat{\beta}_0 \sum_{i}^{n} X_i + \hat{\beta}_1 \sum_{i}^{n} X_i^2 \qquad \ldots\ldots\ldots\ldots (8)$$

وبحل المعادلتين الطبيعيتين الواردة بالعلاقتين (7) و (8) حلاً آنياً أو بالتعويض المباشر، نحصل على القيم التقديرية للمعلمتين $(\hat{\beta}_1 , \beta_0\ )$ على الوجه الآتي :

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} \qquad \ldots\ldots\ldots\ldots\ldots\ldots (9)$$

وبتعويض قيمة $\left(\hat{\beta}_0\right)$ في المعادلة (8)، و بعد التبسيط نحصل على قيمة االتقديرية $\left(\hat{\beta}_1\right)$ كالآتي :

$$\hat{\beta}_1 = \sum_{i}^{n} X_i Y_i - n\,\overline{XY} \qquad \ldots\ldots\ldots\ldots\ldots\ldots (10)$$

إن العلاقة رقم (10) يمكن كتابتها بالشكل الآتي :

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (11)$$

وبتعويض القيم التقديرية للمعلمات $(\hat{\beta}_1 , \beta_0\ )$ في نموذج الانحدار الخطي البسيط

الوارد بالعلاقة رقم (1)، نحصل على معادلة خط الانحدار التنبؤية **(Forecastingrquation)**، و على الوجه الآتي :

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (12)$$

حيث إن :

$\hat{\beta}_0$: تمثل بعد نقطة تقاطع خط الانحدار مع الإحداثي الصادي، عن نقطة الأصل.

$\hat{\beta}_1$: تمثل معامل الانحدار، أو ميل خط الانحدار (**Slope**).

ويعرف معامل الانحدار ($\hat{\beta}_1$)، بأنه :

مؤشر احصائي يفسر مقدار التغير الذي يطرأ على المتغير التابع (X_i) إذا ما تغير المستقل (X_i) بوحدة واحدة.

مثال :

البيانات التالية : تمثل متوسط الدخل الشهري (X) ومتوسط الإنفاق الشهري (Y)، لخمسة عوائل :

متوسط الدخل الشهري (X)	200	320	480	260	620
متوسط الدخل الشهري (Y)	180	240	310	160	400

الطلوب :

(1) تقدير معلمات نموذج الانحدار الخطي البسيط ($\hat{\beta}_0$ ، $\hat{\beta}_1$)

(2) التنبؤ بمتوسط الإنفاق الشهري($\hat{Y}$)لعائلة ما، متوسط دخلها الشهري (780) دينار.

الحل :

(1) تقدير معلمات النموذج($\hat{\beta}_0$ ، $\hat{\beta}_1$)

X	Y	XY	X^2
200	180	36000	40000
320	240	76800	102400
480	310	148800	230400
260	160	41600	67600
620	400	248000	384400
1880 $\overline{X} = 376$	1290 $\overline{Y} = 258$	551200 $\sum XY$	824800 $\sum X^2$

$$\therefore S_{xy} = \sum XY - n\overline{XY}$$
$$= 551200 - 5(376)(258)$$
$$=66160$$
$$S_{xx} = \sum X^2 - n\overline{X}^2$$
$$= 824800 - 5(376)^2$$
$$=117920$$
$$\therefore \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$
$$=\frac{66160}{117920}$$
$$=0.561$$
$$\therefore \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}$$
$$=258-0.561\ (376)$$
$$=47.064$$

وبتعويض القيم التقديرية التقديرية $(\hat{\beta}_0 ، \hat{\beta}_1)$، في نموذج الانحدار الخطي البسيط، سنحصل على النموذج التنبؤي الآتي :

$$\therefore \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$
$$\therefore \hat{Y}_i = 47.064 + 0.561\ X_i \implies Forecastinq\ Model.$$

(2) التنبؤ بمتوسط الانفاق الشهري $(\hat{Y}_i)$ لعائلة دخلها الشهري (780) دينار.

$$\therefore \hat{Y}_i = 47.064 + 0.561\ \ X_i$$
$$\therefore \hat{Y}_i = 47.064 + 0.561^*\ \ (780)$$
$$=484.644 \quad \text{JD.} \quad \text{دج}$$

النتيجة أعلاه، تشير الى أن، العائلة التي دخلها الشهري (780) دينار سيكون متوسط إنفاقها الشهري المتنبأ به $(\hat{Y}_i)$ مساوي الى (484.644) دينار.

3- مؤشرات اختبار جودة توفيق نموذج الانحدار الخطي البسيط:

هناك عدد من المؤشرات الاحصائية، التي يمكن استخدامه لاختبار جودة توفيق نموذج نموذج الانحدار الخطي البسيط، نذكر منها :

1. معامل التحديد (R^2) :Determination Coefficient

يستخدم معامل التحديد (R^2) بشكل عام، لتقدير ما تفسره المتغيرات المستقلة من تغيرات تطرأ على قيم المتغير التابع. ويطلق على معامل التحديد أحياناً بـ (معامل التفسير).

وبناءاً على ما تقدم، يعرف معامل التحديد (R^2)، بإنه : "مؤشر اخصائي يوضح مقدار ما يفسره المتغير المستقل (X_i)من تغير في المتغير التابع (X_i)" و تتراوح قيمة معامل التحديد بين (1.0) أيان$(0\leq R^2 \leq 1)$.

ويمكن إيجاد قيمة معامل التحديد (R^2)، وفقاً للصيغة الآتية :

$$R^2 = \frac{S_{xy}^2}{S_{xx} * S_{yy}} \qquad \dots\dots\dots\dots\dots..(\mathbf{13})$$

2. الخطأ المعياري $(SE_{\hat{y}})$: Standard Error of Estimation

يعرف الخطأ المعياري للتقدير $(SE_{\hat{y}})$، بأنه : "الجذر التربيعي لمجموع مربعات انحرافات القيم الحقيقية (Y_i) عن القيم التقديرية $(\hat{Y}_i$ مقسوماً على (n-2)،" أي إن :

$$SE_{\hat{y}} = \frac{\sqrt{\sum_{i}^{n}(Y_i - \hat{Y}_i)^2}}{\text{n-2}} \Rightarrow \left(\text{الصيغة التعريفية}\right) \dots\dots\dots\dots(\mathbf{14})$$

وبتبسيط صيغة التعريف السابقة، يمكن الحصول على صيغة بديلة لاحتساب الخطأ المعياري

للتقدير $(SE^{\frown})$، التي تعد أكثر سهولة من الناحية التطبيقية، وهي :

$$SE_{\hat{y}} = \frac{\sqrt{\sum^{n} Y_i^2 - \hat{\beta}_0 \sum^{n} Y_i - \hat{\beta}_1 \sum^{n} X_I\, Y_I}}{n-2} \qquad \dots\dots\dots\dots(\mathbf{15})$$

3. اختبار (F) :F-Test

يستخدم اختبار (F) للتحقق من معنوية نموذج الانحدار الخطي البسيط ككل من عدم معنويته ولتحقيق هذا الغرض لابد من اختبار الفرضيتين الآتيتين :

H_0: *The model doesn't significant.*

H_1: *The model is significant.*

إن إحصاءة الاختبار الملائمة للفرضية أعلاه، تأخذ الشكل الآتي :

$$F_{cal.} = \frac{R^2/k-1}{(1-R^2)/n-k} \sim F_{[k-1,n-k]} \quad \ldots\ldots\ldots\ldots.(\mathbf{16})$$

حيث إن :

R^2 : يمثل معامل التحديد.

k : تمثل عدد المتغيرات في النموذج.

n : تمثل عدد أزواج قيم المتغيرين (X) و (Y).

القرار الاحصائي :

أ- يتم قبول فرضية العدم (H_0)، عندما تكون إحصاءة الاختبار $(F_{cal.})$ المحسوبة أقل من القيمة الجدولية، أي إن $[F_{cal.} < F_{tab.}]$، مما يدل ذلك على عدم معنوية نموذج الانحدار الخطي البسيط ككل.

ب-يتم رفض فرضية العدم (H_0)، عندما تكون إحصاءة الاختبار $(F_{cal.})$المحسوبة أكبر من أو تساوي القيمة الجدولية، أي إن $[F_{cal.} \geq F_{tab.}]$، مما يدل ذلك على عدم معنوية نموذج الانحدار الخطي البسيط ككل.

4- اختبار (t) :t-Test

يستخدم اختبار (t) للتحقق من معنوية معلمات نموذج الانحدار الخطي البسيط $(\hat{\beta}_0$ ، $\hat{\beta}_1)$ كل على انفراد، وعلى النحو الآتي :

اختبار معنوية معامل التقاطع $(\hat{\beta}_0)$

للتحقق من معنوية معامل التقاطع $(\hat{\beta}_0)$ الخاص بنموذج الانحدار الخطي البسيط، لابد من اختبار الفرضيتين الآتيتين :

$$H_0: \beta_0 = 0$$
$$H_1: \beta_0 \neq 0$$

إن إحصاءة الاختبار للفرضية أعلاه، تأخذ الشكل التالي :

$$t_{cal.} = \frac{\hat{\beta}_0 - \beta_0}{SE_{\hat{\beta}_0}} \sim t_{(n-2)} \quad \ldots\ldots\ldots\ldots\ldots. (\mathbf{17})$$

حيث إن :

$\hat{\beta}_0$: تمثل القيمة التقديرية للمعلمة (β_0).

$SE_{\hat{\beta}_0}$: تمثل الخطأ المعياري للمعلمة المقدرة $(\hat{\beta}_0)$.

ويمكن إيجاد الخطأ المعياري للمعلمة المقدرة $(\hat{\beta}_0)$ وفقاً للصيغة الآتية :

$$SE_{\hat{\beta}_0} = SE_{\hat{y}} * \sqrt{\frac{1}{n} + \frac{\overline{X}^2}{S_{xx}}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (18)$$

القرار الاحصائي :

أ- يتم قبول فرضية العدم (H_0)، عندما تكون القيمة المطلقة لإحصاءة الاختبار المحسوبة $(|t_{cal.}|)$ أقل من القيمة الجدولية، أي إن $[|t_{cal.}| < t_{(n-2)}]$، مما يدل ذلك على عدم معنوية معامل التقاطع $(\hat{\beta}_0)$

ب- يتم رفض فرضية العدم (H_0)، عندما تكون القيمة المطلقة لاحصاءة الاختبار المحسوبة $(|t_{cal.}|)$ أكبر من أو تساو القيمة الجدولية، أي إن $[|t_{cal.}| \geq t_{(n-2)}]$، مما يدل ذلك على عدم معنوية معامل التقاطع $(\hat{\beta}_0)$ بمعنى إن معامل التقاطع لم يكن مساو للصفر، أي إن (معادلة خط الانحدار لم تمر بنقطة الأصل).

(ب) إختبار معنوية معامل الانحدار $(\hat{\beta}_1)$

للتحقق من معنوية معامل الانحدار $(\hat{\beta}_1)$ بنموذج الانحدار الخطي البسيط، والذي يسمى خط الانحدار (Slope)، لا بد من اختبار الفرضيتين الآتيتين :

$$H_0: \beta_1 = 0$$
$$H_1: \beta_1 \neq 0$$

إن إحصاءة الاختبار للفرضية أعلاه، تأخذ الشكل التالي :

$$t_{cal.} = \frac{\hat{\beta}_1 - \beta_1}{SE_{\hat{\beta}_1}} \sim t_{(n-2)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots. (\mathbf{19})$$

حيث إن :

$\hat{\beta}_1$: تمثل القيمة التقديرية للمعلمة (β_1).

$SE_{\hat{\beta}_1}$: تمثل الخطأ المعياري للمعلمة المقدرة $(\hat{\beta}_1)$.

ويمكن إيجاد الخطأ المعياري للمعلمة المقدرة $(\hat{\beta}_1)$ بالصيغة الآتية :

$$SE_{\hat{\beta}_1} = SE_{\hat{y}} * \sqrt{\frac{1}{S_{xx}}} \qquad \text{.........................} \ (20)$$

القرار الاحصائي :

أ- يتم قبول فرضية العدم (H_0)، عندما تكون القيمة المطلقة لإحصاءة الاختبار المحسوبة $(|t_{cal.}|)$ أقل من القيمة الجدولية، أي إن $[|t_{cal.}| < t_{(n-2)}]$، مما يدل ذلك على عدم معنوية معامل الانحدار $(\hat{\beta}_1)$

ب-يتم رفض فرضية العدم (H_0)، عندما تكون القيمة المطلقة لاحصاءة الاختبار المحسوبة $(|t_{cal.}|)$ أكبر من أو تساو القيمة الجدولية، أي إن $[|t_{cal.}| \geq t_{(n-2)}]$، مما يدل ذلك على عدم معنوية معامل الانحدار $(\hat{\beta}_1)$ نى إن معامل الانحدار لا يساوي صفراً.

مثال :

البيانات التالية : تمثل مصروف الجيب الأسبوعي (X) لثمانية طلاب وعلاماتهم الفصلية (Y) في مساق الرياضيات :

مصروف الطالب (X)	2	4	10	8	2	13	5	12
علامته الفصلية (Y)	97	90	60	55	92	44	82	40

الطلوب :

(1) تقدير معلمات نموذج الانحدار الخطي البسيط ($\hat{\beta}_0$ ، $\hat{\beta}_1$)

(2) حساب معامل التحديد (R^2)، مفسراً النتيجة.

(3) إيجاد قيمة الخطأ المعياري للتقدير $(SE_{\hat{y}})$

(4) اختبار معنوية نموذج الانحدار، مستخدماً $(\propto = 0.01)$.

(5) اختبار معنوية معلمات النموذج ($\hat{\beta}_0$ ، $\hat{\beta}_1$)، مستخدما$(\propto = 0.05)$.

(6) تقدير علامة الطلاب $(\hat{Y}_i)$ إذا كان مصروف الأسبوعي (6) دنانير.

(7)

الحل :

(1) تقدير معلمات النموذج ($\hat{\beta}_0$ ، $\hat{\beta}_1$).

X	Y	XY	X^2	Y^2
2	97	194	4	9409
4	90	360	16	8100
10	60	600	100	3600
8	55	440	64	3025
2	92	184	4	8464
13	44	572	169	1936
5	82	410	25	6724
12	40	482	144	1600
56 $\overline{X} = 7$	560 $\overline{Y} = 70$	3240 $\sum XY$	526 $\sum X^2$	42858 $\sum Y^2$

$$\therefore S_{xy} = \sum XY - n\overline{XY}$$

$$= 3240 - 8(7)(70)$$

$$=-680$$

$$S_{xx} = \sum X^2 - n\overline{X}^2$$

$$= 526 - 8(7)^2$$

$$=134$$

$$\therefore \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

$$=\frac{-680}{134}$$

$$\approx -5.075$$

$$\therefore \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}$$

$$=70 - (-5.075)(7)$$

$$=105.525$$

عليه يكون النموذج التنبؤي (التقديري)، على النحو الآتي :

$$\therefore \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

$$\therefore \hat{Y}_i = 105.525 - 5.075\, X_i \;\; \Longrightarrow Forecasting\ Model.$$

(2) حساب معامل التحديد (R^2) :

$$\because R^2 = \frac{S_{xy}^2}{S_{xx} * S_{yy}}$$

نقوم أولا بإيجاد (S_{yy})، كالآتي :

$$S_{yy} = \sum Y^2 - n\bar{Y}^2$$

$$= 42858 - 8(70)^2$$

$$=3658$$

$$S_{xx} = \sum X^2 - n\bar{X}^2$$

$$= 526 - 8(7)^2$$

$$=134$$

$$\therefore R^2 = \frac{(-680)^2}{134 * 3658}$$

$$=0.943$$

يتضح من خلال قيمة معامل التحديد (R^2) البالغة (0.943)، بأن المتغير المستقل المتمثل بمصروف الطالب الأسبوعي (X)، يفسر ما نسبته (94.3%)من التغيرات التي تطرأ على المتغير التابع المتمثل بعلامة الطالب الفصلية (Y) في مساق الرياضيات، أما النسبةالمتبقية والبالغة (5.7%) فإنها تعود الى متغيرات أخرى لم تدخل في نموذج الانحدار الخطي البسيط قيد الدراسة.

وبناءاً على ما تقدم، يتضح بأن قيمة معامل التحديد (R^2) قريبة جداً من الواحد الصحيح، وهذه النتيجة تدل على جودة توفيق نموذج الانحدار البسيط لبيانات الظاهرة المدروسة، وإن نموذج الانحدار قد مثل الظاهرة أفضل تمثيل.

(3) إيجاد الخطأ المعياري للتقدير($SE_{\hat{Y}}$) :

يمكن إيجاد قيمة الخطأ المعياري لتقدير ($SE_{\hat{Y}}$) وفقاً لإحدى الطريقين الآتيتين :

(أ) طريقة البواقي (الصيغة التعريفية):

$$\because \hat{Y}_i = 105.525 - 5.075\, Y_i$$

X_i	Y_i	$\hat{Y}_i$	$e_i^2 = Y_i - \hat{Y}_i$	$e_i^2 = (Y_i - \hat{Y}_i)^2$
2	97	95.375	1.625	2.6406
4	90	85.225	4.775	22.8006
10	60	54.775	5.225	27.3006
8	55	64.925	-9.925	98.5056
2	92	95.375	-3.375	11.3906
13	44	39.550	4.450	19.8025
5	82	80.150	1.850	3.4225
12	40	44.625	-4.625	21.3906
-	-	-	Zero	207.2536
				$\sum_i^n (Y_i - \hat{Y}_i)^2$

$$SE_{\hat{y}} = \sqrt{\frac{\sum_i^n (Y_i - \hat{Y}_i)^2}{n-2}}$$

$$= \sqrt{\frac{207.2536}{8-2}}$$

$$= \sqrt{34.5423}$$

$$= 5.877$$

(ب) طريقة القيم الأصلية الصيغة البديلة :

$$\therefore \sum Y_i = 560 \quad , \sum Y_i^2 = 42858 \quad , n = 8$$

$$\sum X_i Y_i = 3240 \quad , \hat{\beta}_0 = 105.525 \quad , \hat{\beta}_1 = -5.075$$

$$\therefore SE_{\hat{y}} = \sqrt{\frac{\sum_i^n Y_i^2 - \hat{\beta}_0 \sum_i^n Y_i - \hat{\beta}_1 \sum_i^n X_i Y_i}{n-2}}$$

$$= \sqrt{\frac{42858 - (105.525)(560) - (-5.075)(3240)}{8-2}}$$

$$= \sqrt{\frac{42858 - 59094 + 16443}{6}}$$

$$= \sqrt{\frac{207}{6}}$$

$$= \sqrt{34.5}$$

$$= 5.874$$

تُعد قيمة الخطأ المعياري للتقدير $(SE_{\hat{y}})$ معلمات بيانات المحسوبة لبيانات الظاهرة المدروسة صغيرة، مما يدل على جودة تقدير معلمات نموذج الانحدار، التي تجعل النموذج يتمتع بكفاءة عالية لأغراض عملية التنبؤ بقيم الظاهرة.

(4) اختبار معنوية نموذج الانحدار:

للتحقق من معنوية نموذج الانحدار الخطي البسيط، نقوم باختبار الفرضيتين الآتيتين :

$$H_0: The\ model\ doesn't\ significant.$$
$$H_1: The\ model\ is\ significant.$$

يتم حساب إحصاءة الاختبار (F)، وفقاً للصيغة الآتية :

$$F_{cal.} = \frac{R^2/k-1}{(1-R^2)/n-k}$$

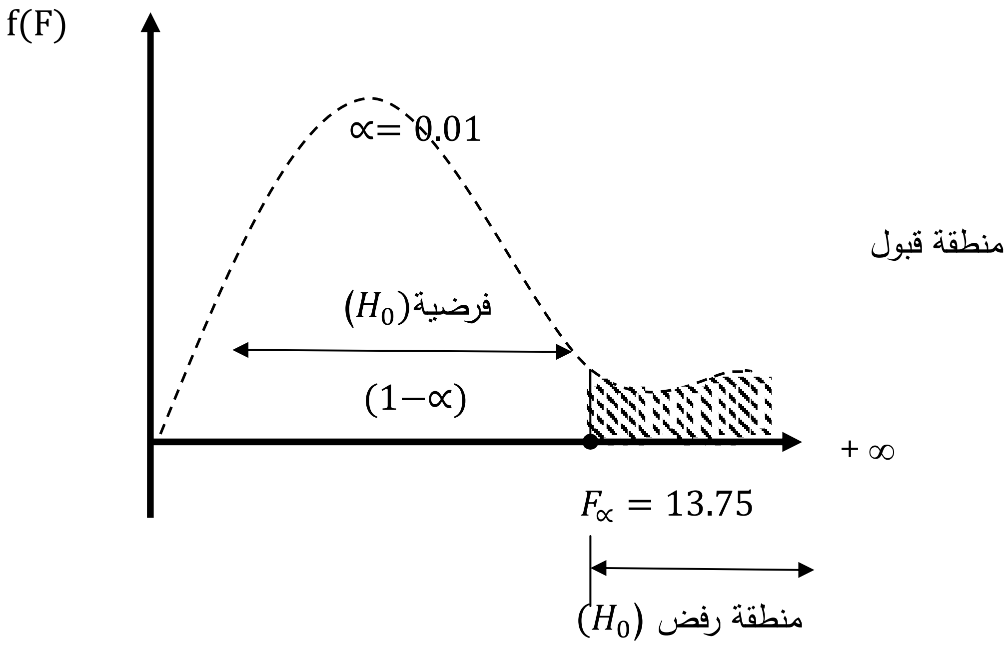
$$\because \mathrm{R}^2 = 0.943 \quad ,\mathrm{k} = 2 \quad ,\mathrm{n} = 8.$$

$$F_{cal.} = \frac{0.943/(2-1)}{(1-0.943)\ \ /8-2}$$

$$= \frac{0.943}{0.0095}$$

$$= 99.263$$

من جداول توزيع (F) بدرجتي حرية (6،1)، عند مستوى المعنوية $(\propto = 0.01)$، نحصل على قيمة (F) الجدولية البالغة (13.75)، و الشكل التالي، يوضح منطقة قبول ورفض فرضية العدم (H_0).



القرار الاحصائي :

بما إن قيمة (F) المحسوبة البالغة (99.263)، هي أكبر من قيمة (F) الجدولية البالغة $F_\propto =$ (13.75) وهذا يعني رفض فرضية العدم (H_0)، مما يدل ذلك على معنوية نموذج الانحدار الخطي البسيط، و بالتالي فإن النموذج يمثل العلاقة بين المتغيرين (X) و (Y) أفضل تمثيل :

(5) اختبار معنوية معلمات النموذج:

أ- اختبار معنوية معامل التقاطع $(\hat{\beta}_0)$

للتحقق من معنوية معامل التقاطع $(\hat{\beta}_0)$، نقوم باختبار الفرضيتين الآتيتين :

$$H_0: \beta_0 = 0$$
$$H_1: \beta_0 \neq 0$$

قبل البدء بحساب إحصاءة الاختبار (t)، نقوم أولاً بحساب $(SE_{\hat{\beta}0})$، كالآتي :

$$\because SE_{\hat{y}} = 5.874 \quad , \overline{X} = 7 \quad n = 8 \ , S_{xx} = 134.$$

$$\therefore SE_{\hat{\beta}0} = SE_{\hat{y}} * \sqrt{\frac{1}{n} + \frac{\overline{X}^2}{S_{xx}}}$$

$$= 5.874 * \sqrt{\frac{1}{8} + \frac{(7)^2}{134}}$$

$$=5.874*(0.7)$$
$$=4.112$$

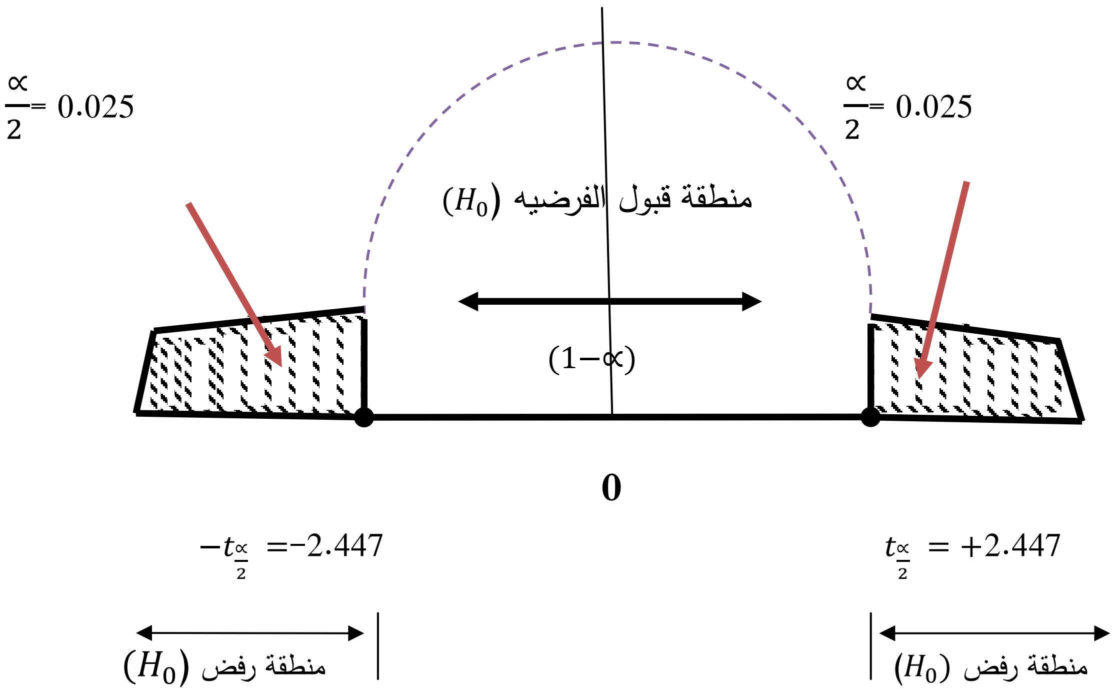
عليه تكون إحصاءة الاختبار، على النحو الآتي :

$$\therefore t_{cal.} = \frac{\hat{\beta}_0 - \beta_0}{SE_{\hat{\beta}0}}$$

$$= \frac{105.525 - 0}{4.112}$$

$$= 25.663$$

من جداول توزيع (t)، بدرجة حرية (6)، عند مستوى المعنوية $(\propto = 0.05)$، نحصل على قيم الجدولية البالغة $\left(t_{\frac{\propto}{2}} = 2.447\right)$ و $\left(-t_{\frac{\propto}{2}} = -2.447\right)$، و الشكل التالي يوضح مناطق قبول ورفض فرضية العدم (H_0).



القرار الاحصائي :

بما إن القيمة المطلقة لاحصاءة الاختبار المحسوبة $(|t_{cal.}|)$ والبالغة (25.663)، هي أكبر من القيمة الجدولية البالغة (′ $(\hat{B}_0)$: $t_{\frac{\propto}{2}}$)، وهذا يعني رفض فرضية العدم $(_0)$ مما يدل ذلك على أن معنوية معامل التقاطعلنموذج الانحدار الخطي البسيط، أي إن معادلة نموذج الانحدار لا تمر بنقطة الأصل.

ب- اختبار معنوية معامل الانحدار $(\hat{\beta}_1)$ باختبار الفرضيتين الآتيتين :

$$H_0: \beta_1 = 0$$
$$H_1: \beta_1 \neq 0$$

نقوم أولاً بحساب $(SE\hat{\ }_1)$ ، وفقاً للصيغة الآتية :

$$SE_{\hat{\beta}1} = SE_{\hat{y}} * \sqrt{\frac{1}{S_{xx}}}$$

$$= 5.874 * \sqrt{\frac{1}{134}}$$

$$=5.874 *(0.086)$$

$$=0.505$$

عليه يمكن حساب إحصاءة الاختبار (t)، وفقاً للصيغة الآتية :

$$t_{cal.} = \frac{\hat{\beta}_1 - \beta_1}{SE_{\hat{\beta}_1}}$$

$$\therefore t_{cal.} = \frac{(-5.075) - 0}{0.505}$$

$$= -10.05$$

من جداول توزيع (t)، بدرجة حرية (6)، عند مستوى المعنوية $(\propto = 0.05)$، نحصل على قيم (t) الجدولية البالغة $\left(t_{\frac{\propto}{2}} = 2.447\right)$ و $\left(-t_{\frac{\propto}{2}} = -2.447\right)$، و الشكل السابق، يوضح مناطق قبول ورفض فرضية العدم (H_0) .

القرار الاحصائي :

بما إن القيمة المطلقة لاحصاءة الاختبار المحسوبة $(|t_{cal.}|)$والبالغة (10.05)، هي أكبر من القيمة الجدولية البالغة $\left(t_{\frac{\propto}{2}} = 2.447\right)$، وهذا يعني رفض فرضية العدم (H_0) مما يدل ذلك على

معنوية معامل الانحدار $(\boldsymbol{\hat{\beta}_1})$ لنموذج الانحدار الخطي البسيط، أي إن قيمة معامل الانحدار لا تساوي صفراً

(6)تقدير علامة الطالب$(\hat{Y}_i)$، إذا كان مصروفه الاسبوعي (6)دنانير :

$$\because \hat{Y}_i = 105.525 - 5.075\ X_i$$

$$\therefore \hat{Y}_i = 105.525 - 5.075 * (6)$$

$$=75.075$$

$$\approx 75$$

من النتيجة أعلاه، يتضح بإن الطالب الذي مصروفه الأسبوعي (6) دنانير ستكون علامته الفصلية التقديرية $(\widehat{Y}_i)$، مساوية الى (75).

5- العلاقة بين معامل الانحدار$(^\frown{}_1)$ومعامل الارتباط (r_p):

The Relationship between the coefficient $(\widehat{\beta}_1)$ and the correlation coefficient (r_p)

يمكن الحصول على معامل الانحدار $(\hat{\beta}_1)$ لنموذج الانحدار الخطي البسيط، بعد معرفة معامل الارتباط البسيط لبيرسون (r_p) بين المتغيرين (X) و (Y)، وذلك من خلال الصيغة الآتية :

$$\hat{\beta}_1 = r_p * \sqrt{\frac{Var(Y)}{Var(X)}} \qquad \text{.............}(21)$$

حيث إن :

Var (X): يمثل تباين المتغير (X)

Var (Y): يمثل تباين المتغير (Y)

من جانب آخر، يمكن الحصول على معامل الارتباط البسيط لبيرسون (r_p) بين المتغيرين (X) و (Y)، بعد معرفة معامل الانحدار $(\hat{\beta}_1)$ لنموذج الانحدار الخطي البسيط وذلك من خلال الصيغة الآتية :

$$r_p = \hat{\beta}_1 * \sqrt{\frac{Var(X)}{Var(Y)}} \qquad \text{.............}(22)$$

مثال :

إذا كان لديك المعلومات الآتية :

$$\sum_{i=1}^{10}(X_i - 5)^2 = 105 \quad , \sum_{i=1}^{10}(Y_i - 1)^2 = 42 \quad , r_p = 0.8$$

المطلوب :

(1)تقدير معلمات نموذج الانحدار الخطي البسيط. $(\hat{\beta}_1 ، \hat{\beta}_0)$.

(2) حساب معامل التحديد (R^2)مفسراً النتيجة.

(3) التنبؤ بقيمة $(\widehat{Y}_i)$، إذا كانت قيمة $(X_i = 15)$.

الحل :

(1)تقدير معلمات النموذج $(\hat{\beta}_1 ، \hat{\beta}_0)$

$$\because \hat{\beta}_1 = r_p * \sqrt{\frac{Var(Y)}{Var(X)}}$$

نقوم بحساب تباين المتغيرين (X) و (Y)كالآتي :

$$Var(X) = \frac{\sum_{i}^{10}(X_i - 5)^2}{n-1}$$

$$= \frac{105}{10-1}$$
$$=11.667$$

$$Var(Y) = \frac{\sum_{i}^{10}(Y_i - 1)^2}{n-1}$$
$$= \frac{42}{10-1}$$

$$= 4.667$$

$$\because \hat{\beta}_1 = 0.8 * \sqrt{\frac{4.667}{11.667}}$$
$$= 0.8 * (0.632)$$
$$=0.506$$

$$\therefore \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$
$$=1-(0.506)\,(5)$$

$$=1-2.53$$
$$= -1.53$$

عليه يكون النموذج التقديري (التنبؤي) على النحو الآتي :

$$\because \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$
$$\therefore \hat{Y}_i = -1.53 + 0.506\, X_i \implies Forecasting\ Model.$$

(1) حساب معامل التحديد (R^2) :

يعرف التحديد (R^2)، بأنه :"مربع قيمة معامل الارتباط الخطي البسيط" أي إن :

$$R^2 = r_p^2$$
$$= (0.8)^2$$
$$= 0.64$$

يتضح من خلال قيمة معامل التحديد (R^2)، البالغة '0.64)، بأن المتغير المستقل (X)، يفسر ما نسبته (64%) من التغيرات التي تطرأ على المتغير التابع (Y)، أما النسبة المتبقية و البالغة (36%) فإنها تعود الى متغيرات أخرى لم تدخل في نموذج الانحدار الخطي البسيط قيد الدراسة.

(2) التنبؤ بقيمة $(\hat{Y}_i)$، إذا كانت قيمة $(X_i = 15)$:

$$\because \hat{Y}_i = 1.53 + 0.506\, X_i$$

$$\therefore \hat{Y}_i = -1.53 + 0.506 * (15)$$

$$= -1.53 + 7.59$$

$$= 6.06$$

6- الانحدار الخطي المتعدد Multiple Linear Regression:

يعرف الانحدار الخطي المتعدد بأنه : "عملية تقدير العلاقة الخطية بين عدة متغيرات، يٌعد أحدها متغيراً تابعاً **(Dependent Variable)** والمتغيرات الأخرى تٌعد متغيرات مستقل **(Independent Variable)**."

وستنصب دراستنا على دراسة (3) ثلاثة متغيرات، أحدها متغير تابع (Y_i)، و متغيرين مستقلين فقط هما (X_1) و (X_2)، كما موضحة في نموذج الانحدار المتعدد الآتي :

$$Y_i = a + b_i X_{1i} + b_2 X_{2i} + u_i \qquad \ldots\ldots\ldots\ldots\ldots\ldots..(\mathbf{20})$$

من المعادلة رقم (20) نحصل على مجموع مربعات الأخطاءكالآتي $\left(\sum_i^n u_i^2\right)$

$$\sum_i^n u_i^2 = \sum_i^n (Y_i - \hat{a} - \hat{b}_i X_{1i} - \hat{b}_2 X_{2i})^2 \qquad \ldots\ldots\ldots\ldots\ldots..(\mathbf{21})$$

بتطبيق طريقة المربعات الصغرى الاعتيادية (OLS) التي تجعل من مجموع مربعات الأخطاء أقل ما يمكن، من خلال إجراء التفاضل الجزئي بالنسبة للمعلمات $(\hat{b}_2 , \hat{b}_1, \hat{a}\ )$ بعد مساواتها بالصفر، حصلنا على المعادلات الطبيعية **(Normal Equations)** الآتية :

$$\sum_i^n Y_i = n\,\hat{a} + \hat{b}_1 \sum_i^n X_{1i} + \hat{b}_2 \sum_i^n X_{2i} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(\mathbf{22})$$

$$\sum_i^n X_{1i}\, Y_i = \hat{a} \sum_i^n X_{1i} + \hat{b}_1 \sum_i^n X_{1i}^2 + \hat{b}_2 \sum_i^n X_{1i}\, X_{2i} \quad \ldots\ldots\ldots(\mathbf{23})$$

$$\sum_i^n Y_{2i}\, Y_i = \hat{a} \sum_i^n X_{2i} + \hat{b}_1 \sum_i^n X_{1i}\, X_{2i} + \hat{b}_2 \sum_i^n X_{2i}^2 \qquad \ldots\ldots\ldots(\mathbf{24})$$

من المعادلة رقم (22) محصل على :

$$\overline{Y} = \hat{a} + \hat{b}_1 \overline{X}_1 + \hat{b}_2 \overline{X}_2 \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(\mathbf{25})$$

عليه فإن المعلمة $(\hat{a})$ يمكن الحصول عليها من المعادلة رقم (25)، كالآتي :

$$\hat{a} = \overline{Y} - \hat{b}_1 \overline{X}_1 - \hat{b}_2 \overline{X}_2 \qquad \ldots\ldots\ldots\ldots\ldots\ldots.(\mathbf{26})$$

وحيث أن المعادلة التنبؤية**(Forecasting Equation)**هي :

$$\hat{Y}_i = \hat{a} + \hat{b}_1 X_{1i} + \hat{b}_2 X_{2i} \qquad \ldots\ldots\ldots\ldots.(\mathbf{27})$$

بطرح المعادلة (25) من المعادلة رقم(27)، نحصل على صيغة الانحرافات أي إن :

$$\hat{Y}_i - \overline{Y} = \hat{b}_1 \big(X_{1i} - \overline{X}_1\big) + \hat{b}_2 (X_{2i} - \overline{X}_2)$$

أي إن :

$$\hat{y}_i = \hat{b}_1 x_{1i} + \hat{b}_2 x_{2i} \qquad \Longrightarrow Deviation\ Formula$$

عليه فإن الأخطاء التقديرية (e_i) تآتي :

$$\because e_i = y_i - \hat{y}_i$$

$$\therefore \sum_i^n e_i^2 = \sum_i^n (y_i - \hat{y}_i)^2$$

$$\therefore \sum_i^n e_i^2 = \sum_i^n [y_i - \hat{b}_i x_{1i} - \hat{b}_2 x_{2i}]^2 \quad \dots\dots\dots\dots.(\mathbf{28})$$

ولجعل مجموع مربعات الأخطاء $\left(\sum_i^n e_i^2\right)$ يمكن، نقوم بإجراء التفاضل الجزئي على المعادلة (28) بالنسبة الى $(\hat{b}_1)$ و $(\hat{b}_2)$ و مساواتها بالصفر، نحصل على :

$$\sum_i^n x_{1i} y_i = \hat{b}_1 \sum_i^n x_{1i}^2 + \hat{b}_2 \sum_i^n x_{1i}\, x_{2i} \qquad \dots\dots\dots\dots(\mathbf{29})$$

$$\sum_i^n x_{2i} y_i = \hat{b}_1 \sum_i^n x_{1i}\, x_{2i} + \hat{b}_2 \sum_i^n x_{2i}^2 \qquad \dots\dots\dots\dots(\mathbf{30})$$

وللحصول على الصيغ الاحصائية التقديرية الخاصة بإيجاد قيمة $(\hat{b}_1)$ و $(\hat{b}_2)$ نقوم بتطبيق طريقة (كريمر)، كالآتي :

$$\hat{b}_1 = \frac{\begin{vmatrix} \sum_i^n x_{2i} y_i & \sum_i^n x_{2i}^2 \end{vmatrix} \begin{vmatrix} \sum_i^n x_{1i} y_i & \sum_i^n x_{1i} x_{2i} \end{vmatrix}}{\begin{vmatrix} \sum_i^n x_{1i}\, x_{2i} & \sum_i^n x_{2i}^2 \end{vmatrix} \begin{vmatrix} \sum_i^n x_{1i}^2 & \sum_i^n x_{1i} x_{2i} \end{vmatrix}}$$

$$\therefore \hat{b}_1 = \frac{\left[\sum_i^n x_{1i} y_i\right]\left[\sum_i^n x_{2i}^2\right] - \left[\sum_i^n x_{2i} y_i\right]\left[\sum_i^n x_{1i} x_{2i}\right]}{\left[\sum_i^n x_{1i}^2\right]\left[\sum_i^n x_{2i}^2\right]\left[\sum_i^n x_{1i} x_{2i}\right]} \qquad \dots\dots\dots.(31)$$

كذلك يمكن الحصول على :

$$\hat{b}_2 = \frac{\begin{vmatrix} \sum_i^n x_{1i}\, x_{2i} \sum_i^n x_{2i}\, y_i \end{vmatrix} \begin{vmatrix} \sum_i^n x_{1i}^2 \sum_i^n x_{1i} y_i \end{vmatrix}}{\begin{vmatrix} \sum_i^n x_{1i}\, x_{2i} \sum_i^n x_{2i}^2 \end{vmatrix} \begin{vmatrix} \sum_i^n x_{1i}^2 \sum_i^n x_{1i} x_{2i} \end{vmatrix}}$$

$$\therefore \hat{b}_2 =. \frac{\left[\sum_i^n x_{2i} y_i\right]\left[\sum_i^n x_{1i}^2\right] - \left[\sum_i^n x_{1i} y_i\right]\left[\sum_i^n x_{1i} x_{2i}\right]}{\left[\sum_i^n x_{1i}^2\right]\left[\sum_i^n x_{2i}^2\right]\left[\sum_i^n x_{1i} x_{2i}\right]} \quad ....(32)$$

عليه فإن الصيغ الواردة في (26) و (31) و (32) تستخدم في عملية تقدير المعلمات على التوالي : $(\hat{b}_2$ و$\hat{b}_1$ ، $\hat{a})$

وبتعويض القيم التقديرية للمعلمات أعلاه، في نموذج الانحدار الخطي المتعدد الواردة في العلاقة رقم (20)، نحصل على النموذج التنبؤي(**Forecasting Model**)، و كالآتي :

$$\hat{Y}_i = \hat{a} + \hat{b}_1 X_{1i} + \hat{b}_2 X_{2i} \quad \Rightarrow Forecasting\ Model \qquad \dots\dots\dots\dots\dots.\dots.(33)$$

7- المؤشرات الاحصائية المستخدمة في اختبار توفيق نموذج الانحدار المتعدد

أدناه أهم المؤشرات الاحصائية التي تستخدم لاختبار جودة توفيق نموذج الانحدار المتعدد :

1. معامل التحديد (R^2) **Determination Coefficient:**

يمكن ايجاد معامل التحديد (R^2) باستخدام الصيغة الآتية :

$$R^2 = \frac{\hat{b}_1 \sum_i^n x_{1i}\, y_i + \hat{b}_2 \sum_i^n x_{2i}\, y_i}{\sum_i^n {y_1}^2}$$

2. معامل التحديد المعدل (R^2) Adjusted Determination Coefficient:

يمكن ايجاد معامل التحديد المعدل $(\overline{R}^2)$ بموجب الصيغة الآتية :

$$\overline{R}^2 = 1 - \frac{n-1}{n-k} * (1 - R^2)$$

3. الخطأ المعياري للتقدير $(SE_{\hat{y}})$: Standard Error of Estimation

يمكن ايجاد الخطأ المعياري للتقدير $(SE_{\hat{y}})$ بموجب الصيغة الآتية :

$$SE_{\hat{y}} = \frac{\sqrt{\sum_{i}^{n} y_i^2 - \hat{b}_1 \sum_{i}^{n} x_{1i} y_i - \hat{b}_2 \sum_{i}^{n} x_{2i} y_i}}{n-k}$$

4. اختبار (F): F-test

يستخدم اختبار (F) للوقوف على معنوية نموذج الانحدار المتعدد من عدم معنويته، حيث أن معيار الاختبار المستخدم هو :

$$F = \frac{R^2 \ / k-1}{(1-R^2) \ / n-k} \sim F_{(k-1, n-k, \propto)}$$

القرار الاحصائي :

أ- يتم قبول فرضية العدم (H_0)، وذلك عندما تكون إحصاءة الاختبار المحسوبة $(F_{cal.})$ أقل من القيمة الجدولية، $(F_{tab.})$ بدرجتي الحرية $[(n-k), (k-1)]$، و مستوى المعنوية $(\propto)$وهذا يعني إن نموذج الانحدار الخطي المتعدد غير معنوي عند مستوى المعنوية المذكور.

ب- يتم رفض فرضية العدم (H_0)، عندما تكون القيمة المطلقة لاحصاءة الاختبار المحسوبة $(F_{cal.})$أكبر من أو تساو القيمة الجدولية، $(F_{tab.})$أي إن $[F_{cal.} \geq F_{tab.}]$، وهذا يعني إن نموذج الانحدار الخطي المتعدد يعد معنزي عند مستوى المعنوية المطلوب $(\propto)$.

8- اختبار (t): t-test

يستخدم هذا الاختبار للتحقق من معنوية معلمات النموذج كل على انفراد، حيث إن معيار الاختبار المستخدم هو :

$$t = \frac{\hat{b}}{S_{\hat{b}}} \sim t_{(n-k,\propto)}$$

ولنموذج الانحدار الذي يحتوي على (3) معلمات $(\hat{b}_2, \hat{b}_1, \hat{a})$ يكون الخطأ المعياري للمعلمتين $(\hat{b}_2$ و $\hat{b}_1)$ كالآتي :

$$S_{\hat{b}_1} = SE_{\hat{y}}^{*} \frac{\sqrt{\sum_{i}^{n} x_{2i}^2}}{\sqrt{X'X}}$$

$$S_{\hat{b}_2} = SE_{\hat{y}}^{*} \frac{\sqrt{\sum_{i}^{n} x_{21i}^2}}{\sqrt{X'X}}$$

حيث إن :

$$|X'X| = \left(\sum x_{1i}^2\right)\left(\sum x_{2i}^2\right) - \left(\sum x_{1i}x_{2i}\right)^2$$

القرار الاحصائي :

أ- يتم قبول فرضية العدم (H_0)، وذلك عندما تكون القيمة المطلقة لإحصاءة الاختبار $(|t_{cal.}|)$ أقل من القيمة الجدولية، $(t_{tab.})$أي إن$[|t_{cal.}| < t_{tab.}]$ بدرجة حرية $(n-1)$ عند مستوى المعنوية $(\propto)$ وهذا يعني إن المعلمة $(\hat{b})$ غير معنوية عند مستوى المعنوية المذكور.

ب-يتم قبول فرضية العدم (H_0)، وذلك عندما تكون القيمة المطلقة لإحصاءة الاختبار $(|t_{cal.}|)$ أكبر من أو تساوي القيمة الجدولية، $(t_{tab.})$أي إن$[|t_{cal.}| \geq t_{tab.}]$ بدرجة

حرية $(n-k)$ عند مستوى المعنوية $(\propto)$ وهذا يعني إن المعلمة $(\hat{b})$ تعد معنوية معنوية عند مستوى المعنوية المذكور.

مثال :

البيانات التالية، تمثل كميات الطلب (Y_i) على سلعة معينة لعشرة (10) أسر، وسعر الوحدة الواحدة (X_{1i}) من السلعة، و الدخل الشهري (X_{2i})للأسرة :

كميات الطلب (Y_i)	3	5	6	4	6	4	9	8	9	6
سعر الوحدة(X_{1i})	11	8	7	8	6	9	5	5	4	7
الدخل الشهري (X_{2i})	100	120	130	100	120	110	140	130	140	110

الطلوب :

1) تقدير معلمات نموذج الانحدار الخطي المتعدد $(Y_i = b_0 + b_1X_{1i} + b_2 + X_{2i})$

2) حساب معامل التحديد (R^2)، وعامل التحديد المعدل $(\overline{R}^2)$ تائج.

3) إيجاد قيمة الخطأ المعياري للتقدير $(SE_{\hat{y}})$

4) اختبار معنوية نموذج الانحدارالخطي المتعدد، عند مستوى المعنوية ،$(\propto= 0.01)$.

5) اختبار معنوية معلمات النموذج (1 ،$\hat{b}_2$)، عند مستوى المعنوية$(\propto= 0.05)$.

6) التنبؤ بكمية الطلب $(\hat{Y}_i)$ $[X_{2i} = 200\ ,\ X_{1i} = 1$

الحل :

(1) تقدير معلمات النموذج $(\hat{b}_{2i}, \hat{b}_{1i}, \hat{b}_0)$.

Y_i	X_{1i}	X_{2i}	y_i	x_{1i}	x_{2i}	y_i^2	x_{1i}^2	x_{2i}^2	$x_{1i}y_i$	$x_{2i}y_i$	$x_{1i}x_{2i}$
3	11	100	-3	+4	-20	9	16	400	-12	60	-80
5	8	120	-1	+1	0	1	1	0	-1	0	0
6	7	130	0	0	+10	0	0	100	0	0	0
4	8	100	-2	+1	-20	4	1	400	-2	40	-20
6	6	120	0	-1	0	0	1	0	0	0	0
4	9	110	-2	+2	-10	4	4	100	-4	20	-20
9	5	140	3	-2	+20	9	4	400	-6	60	-40
8	5	130	2	-2	+10	4	4	100	-4	20	-20
9	4	140	3	-3	+20	9	9	400	-9	60	-60
6	7	110	0	0	-10	0	0	100	0	0	0
60	70	1200	0	0	0	40	40	2000	-38	260	-240
$\sum_i^n Y_i$	$\sum_i^n X_{1i}$	$\sum_i^n X_{2i}$	$\sum_i^n y_i$	$\sum_i^n x_{2i}$	$\sum_i^n x_{2i}$	$\sum_i^n y_i^2$	$\sum_i^n X_{1i}^2$	$\sum_i^n X_{2i}^2$	$\sum_i^n x_{1i}y_i$	$\sum_i^n \boldsymbol{X_{2i}Y_i}$	$\sum_i^n \boldsymbol{X_{1i}X_{2i}}$

حيث إن :

$$y_i = Y_i - \bar{Y}$$
$$x_{1i} = X_{1i} - \bar{X}_1$$
$$x_{2i} = X_{2i} - \bar{X}_2$$

عليه تكون الأوساط الحسابية للمتغيرات (X_{2i}, X_{1i}, Y_i)، على النحو الآتي :

$$\bar{Y} = 6 \quad , \quad \bar{X}_1 = 7 \quad , \quad \bar{X}_2 = 120$$

$$\because \hat{b}_1 = \frac{\left(\sum_i^n x_{1i}\, y_i\right)\left(\sum_i^n x_{2i}^2\right) - \left(\sum_i^n x_{2i}\, y_i\right)\left(\sum_i^n x_{1i}\, x_{2i}\right)}{\left(\sum_i^n x_{1i}^2\right)\left(\sum_i^n x_{2i}^2\right) - \left(\sum_i^n x_{1i} x_{2i}\right)^2}$$

$$= \frac{-38(2000) - (260)(-240)}{40(2000) - (-240)^2}$$

$$= \frac{-13600}{22400}$$

$$= -0.607$$

$$\hat{b}_2 = \frac{\left(\sum_i^n x_{2i}\, y_i\right)\left(\sum_i^n x_{1i}^2\right) - \left(\sum_i^n x_{1i}\, y_i\right)\left(\sum_i^n x_{1i}\, x_{2i}\right)}{\left(\sum_i^n x_{1i}^2\right)\left(\sum_i^n x_{2i}^2\right) - \left(\sum_i^n x_{1i} x_{2i}\right)^2}$$

$$= \frac{260(40) - (-38)(-240)}{40(2000) - (-240)^2}$$

$$= \frac{1280}{22400}$$

$$= 0.057$$

$$\therefore \hat{b}_0 = \overline{Y} - \hat{b}_1 \overline{X}_1 - \hat{b}_2 \overline{X}_2$$

$$= 6 - (-0.607)(7) - 0.057(120)$$

$$= 3.409$$

$$\because \hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_{1i} + \hat{b}_2 X_{2i}$$

$$\therefore \hat{Y}_i = 3.409 - 0.607X_{1i} + 0.057X_{2i} \implies \textit{Forecasting Model.}$$

(2) معامل التحديد (R^2) و معامل التحديد المعدل ($\overline{R}^2$):

$$\therefore R^2 = \frac{\hat{b}_1 \sum_i^n x_{1i}\, y_i + \hat{b}_2 \sum_i^n x_{2i} y_i}{\sum_i^n y_i^2}$$

$$= \frac{-0.607(-38) + 0.057(260)}{40}$$

$$= 0.95$$

$$\therefore \overline{R}^2 = 1 - \frac{n-1}{n-k} * (1 - R^2)$$

$$= 1 - \frac{10-1}{10-3} - (1 - 0.95)$$

$$= 1 - \frac{9}{7} * (0.05)$$

$$= 1 - 0.06$$

$$= 0.94$$

من أعلاه يتضح بأن كل من سعر الوحدة الواحدة (X_{1i}) و الدخل الشهري (X_{2i}) يفسران ما مقداره (94%) من التغيرات التي تطرأ على الكمية المطلوبة من سلعة ما (Y_i)، أما النسبة المتبقية (6%) فإنها تعود الى متغيرات أخرى غير داخلة في نموذج الانحدار المتعدد.

(3) حساب الخطأ المعياري للتقدير $(SE_{\hat{y}})$

$$\because SE_{\hat{y}} = \frac{\sqrt{\sum_{i}^{n} y_i^2 - \hat{b}_1 \sum_{i}^{n} x_{1i} y_i - \hat{b}_2 \sum_{i}^{n} x_{2i} y_i}}{n-k}$$

$$= \frac{\sqrt{40 - (-0.607)(-38) - 0.057(260)}}{10\text{-}3}$$

$$= \sqrt{\frac{2.118}{7}}$$

$$= \sqrt{0.303}$$

$$= 0.55$$

(5)اختبار معنوية نموذج الانحدار المتعدد :

يتم اختبار معنوية النموذج، من خلال اختبار الفرضيتين الآتيتين :

H_0: *The model doesn't significant.*
H_1: *The model is significant.*

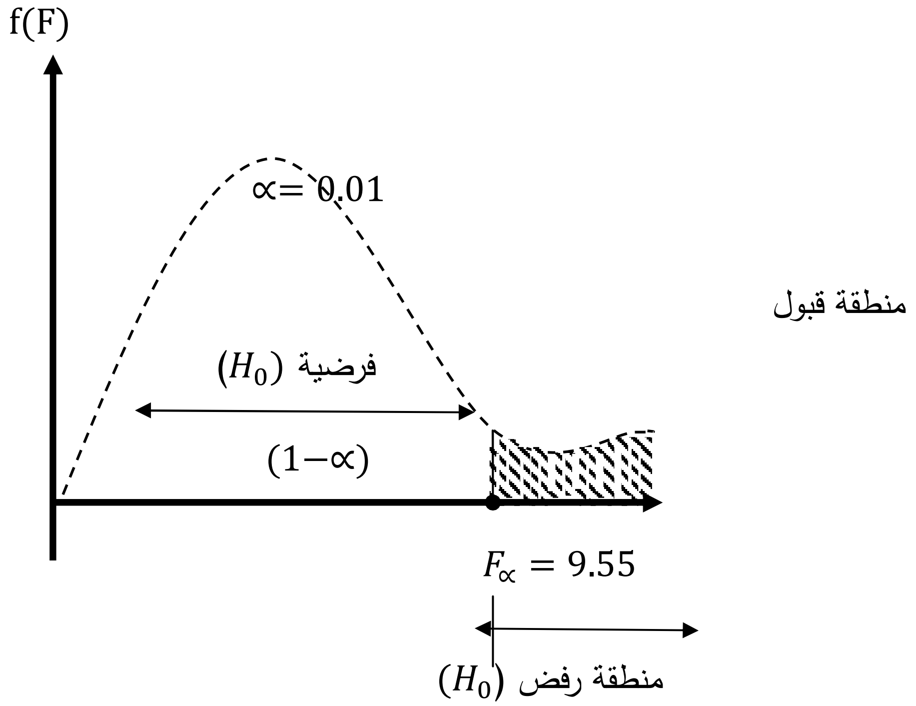
لاختبار الفرضية السابقة، نقوم بحساب إحصاءة الاختبار (F)، وفقاً للصيغة الآتية :

$$F_{cal.} = \frac{R^2/k-1}{(1-R^2)/n-k}$$

$$= \frac{0.95/(3-1)}{(1-0.095)/(10-3)}$$

$$= 66.5$$

من جداول توزيع (F)، بدرجتي حرية (7،2)، عند مستوى المعنوية ($\propto = 0.01$)، نحصل على قيمة (F)، الجدولية البالغة (9.55)، و الشكل التالي يوضح منطقة قبول ورفض فرضية العدم (H_0)



القرار الاحصائي :

بما إن قيمة (F) المحسوبة البالغة (66.5)،هي أكبر من قيمة ($F_\propto$) الجدولية البالغة (9.55) وهذا يعني رفض فرضية العدم (H_0)، مما يدل ذلك على معنوية نموذج الانحدار الخطي المتعددعند مستوى المعنوية ($\propto = 0.01$).

(6) اختبار معنوية معلمات النموذج $[\hat{b}_2 , \hat{b}_1]$:

$$\because |X'X| = \begin{vmatrix} \sum_i^n x_{1i}^2 & \sum_i^n x_{1i}x_{2i} \\ \sum_i^n x_{1i}x_{2i} & \sum_i^n x_{2i}^2 \end{vmatrix}$$

$$= \begin{vmatrix} 40 & -240 \\ -240 & 2000 \end{vmatrix}$$

$$= 40(2000) - (-240)^2$$

$$= 22400$$

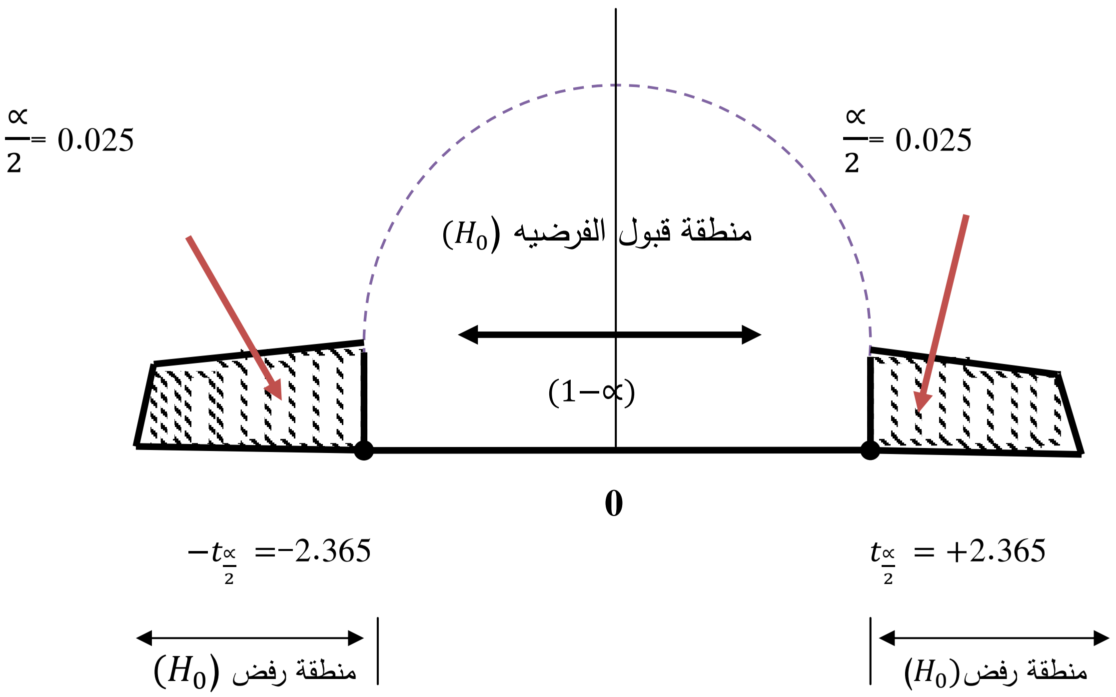
$$\therefore S_{\hat{b}_1} = SE_{\hat{y}} * \sqrt{\frac{\sum_i^n x_{2i}^2}{|X'X|}}$$

$$= 0.55 * \sqrt{\frac{2000}{22400}}$$

$$= 0.55(0.3)$$
$$= 0.165$$

$$\therefore t_{cal.} = \frac{\hat{b}_1}{S_{\hat{b}1}}$$
$$= \frac{-0.607}{0.165}$$
$$= -3.679$$

من جداول توزيع (t) بدرجة حرية (7)، عند مستوى المعنوية $(\propto = 0.05)$، نحصل على قيم (t) الجدولية وهي $(t_{\propto/2} = 2.365)$ و $(-t_{\propto/2} = -2.365)$، و الشكل التالي يوضح مناطق قبول فرضية العدم (H_0).



القرار الاحصائي :

بما إن القيمة المطلقة لإحصاءة الاختبار المحسوبة($|t_{cal.}|$) والبالغة (3.679)، هي أكبر من القيمة الجدولية البالغة $\left(t_{\frac{\alpha}{2}} = 2.365\right)$، وهذا يعني رفض فرضية العدم (H_0). مما يدل على معنوية المعلمة ، عند $(\hat{b}_1 = -0.607)$ $(\alpha = 0)$.

كذلك نقوم بإيجاد $(S_{\hat{b}1})$ كالآتي :

$$S_{\hat{b}_2} = SE_{\hat{y}} * \sqrt{\frac{\sum_i^n x_{2i}^2}{|X'X|}}$$

$$= 0.55 * \sqrt{\frac{40}{22400}}$$

$$= 0.55(0.04)$$

$$= 0.022$$

$$\therefore t_{cal.} = \frac{\hat{b}_2}{S_{\hat{b}2}}$$

$$= \frac{0.057}{0.022}$$

$$= 2.59$$

من جداول توزيع (t) بدرجة حرية (7)، عند مستوى المعنوية ($\propto = 0.05$)، نحصل على قيم (t) الجدولية وهي ($t_{\frac{\propto}{2}} = 2.365$) و ($-t_{\frac{\propto}{2}} = -2.365$)، و الشكل السابق يوضح مناطق قبول فرضية العدم (H_0).

القرار الاحصائي :

بما إن القيمة المطلقة لإحصاءة الاختبار المحسوبة($|t_{cal.}|$)والبالغة (2.59)، هي أكبر من القيمة الجدولية البالغة$\left(t_{\frac{\propto}{2}} = 2.365\right)$، وهذا يعني رفض فرضية العدم (H_0). مما يدل ذلك على معنوية المعلمة عند ه ($\hat{b}_2 = -0.057$) ($\propto =$).

9- التنبؤ بكمية الطلب (Y_i) عند $[X_{2i} = 200 , X_{1i} = 15]$

$$\because \hat{Y}_i = 3.409 - 0.607X_{1i} + 0.057X_{2i}$$

$$\therefore \hat{Y}_i = 3.409 - 0.607(15) + 0.057(200)$$

$$= 5.704$$

$$\hat{Y}_\iota = 6$$

تحليل السلاسل الزمنية: أسهم تطور أساليب تحليل السلاسل الزمنية خلال السنوات الأخيرة في تحقيق طرق دقيقة للتنبؤ والحصول من خلالها على نتائج تساعد على اتخاذ قرارات سليمةوتؤدي إلى تحليل سليم للمتغيرات والعلاقات الاقتصادية وبذلك يمكن تجنب الآثار العكسية لتحليل السلسلة الزمنية بطرق غير دقيقة.

1 مفهوم السلسلة الزمنية :

هي قيم ظاهرة من الظواهر في سلسلة تواريخ متعاقبة سواء كانت هذه أياما، أشهر، أو سنوات ..الخ

تعرف أيضا بأنها عبارة عن مجموعة من المشاهدات مأخوذة على فترات زمنية وذلك نتيجة تعقب هذه الظاهرة لفترة زمنية طويلة نسبيا وبصفة متتابعة وفي الغالب هذه الفترة الزمنية منتظمة.وهي بذلك تحتوي على متغيرين أحدهما مستقل وهو الزمن والآخر تابع وهو قيمة الظاهرة.

من خلال التعاريف يتضح لنا أن السلسة الزمنية هي قيم لظاهرة ما متسلسلة حسب الزمن تبين تطور هذه الظاهرة وتقوم على تفسير المتغير التابع الزمن أو بسلوك نفس المتغير في الماضي.

تتكون أي سلسلة زمنية من قيم متتالية لمتغير إحصائي خلال مدة زمنية منتظمة، أسبوع، فصل أو سنة وتتبع قيمة أي متغير إحصائي في تغيرها الزمن،وتخضع في تغييرها لفترة طويلة من الوقت لعدة مؤثرات بعضها منتظم يمكن ترتيبها ومعرفة درجة تأثيرها وبعضها غير منتظم عشوائي هذه المؤثرات تسمى مركبات السلسلة الزمنية، لتعتبر بذلك علاقة دالية بين ظاهرة معينة(y) و الزمن (t).

$$Y = f(t)$$

2 مركبات السلسلة الزمنية:

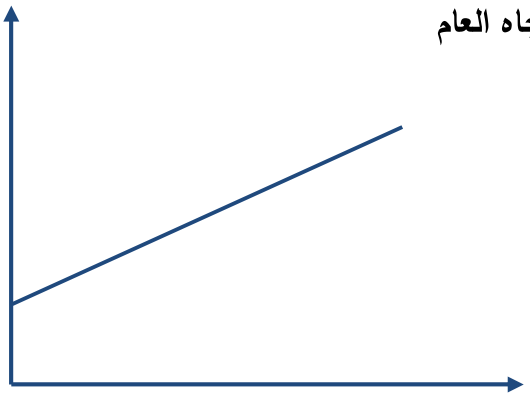
تقوم دراسة السلاسل الزمنية على تحليلها إلى مركباتها أو العناصر المكونة لها لعزلها ومعرفة مدى تأثير كل منها على الظاهرة المشاهدة، و معرفة تطورها عبر الزمن، وبذلك يكون القصد من التحليل رد القيمة الكلية للظاهرة إلى عناصرها المكونة لها ،حتى يصبح بالإمكان القيام بالتقديرات اللازمة والتنبؤات الضرورية، و تكون هذه المركبات منتظمة و غير منتظمة.

المركبات المنتظمة:و يتكرر وجودها في السلسلة، و في مواضع محدّدة، و هي ثلاث:

أ) مركبة الاتجاه العام (التغيرات الموسمية)

وهي أساسية في حركة السلسة الزمنية، تتميز بالنمو الطبيعي المضطرب أو التقلص الطبيعي المتدرج للظاهرة المشاهدة ،وتنعدم ملاحظة هذه التغيرات في فترة قصيرة (تغيرات تحدث ببطء) تأخذ شكلها تدريجيا فتستغرق وقت طويل مما يكسبها صفة الديمومة والاستمرار.

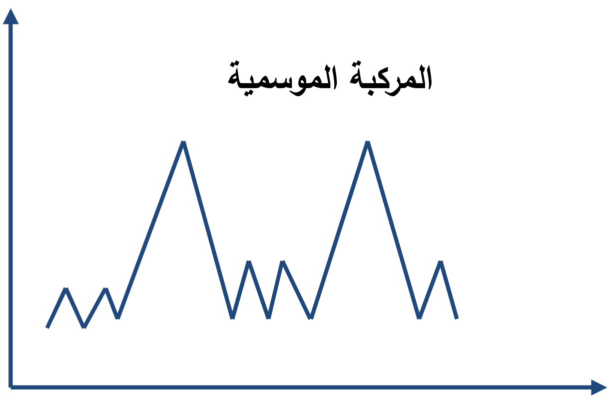
مركبة الاتجاه العام تعبر عن تطور الظاهرة عبر الزمن ولا تكون واضحة إلا في الفترات الطويلة وتوضح لنا اتجاه الظاهرة ،ويرمز لها بالرمز T .



.

ب)المركبة الموسمية (التغيرات الموسمية)

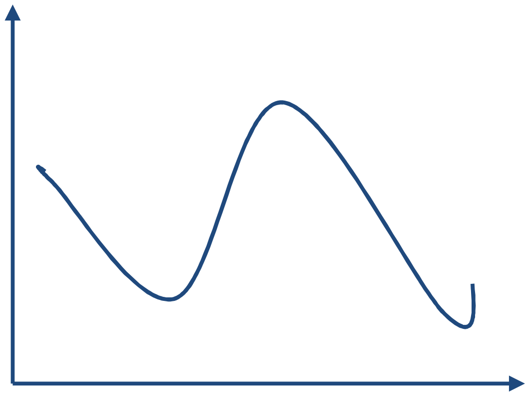
التغيرات الموسمية هي التغيرات التي تحدث بانتظام في وحدات زمنية متعاقبة كشهر معين من السنة أو يوم معين،أو ساعة معينة،أو هي عبارة عن تقلبات تتكرر على نفس الوتيرة كل سنة ويرمز لها ب S.



ج) المركبة الدورية (التغيرات الدورية)

تنعكس هذه المركبة في السلاسل الزمنية طويلة الأجل والتي تبرز أثر انتقال الأحوال الاقتصادية مثلا من الكساد إلى الإنتعاش، فالرواج ثم الركود وهكذا دواليك، وهي تشبه التغيرات الموسمية إلا أنها تتم في فترات أطول نسبيا من الفترات الموسمية،ويرمز لها بالرمز C.

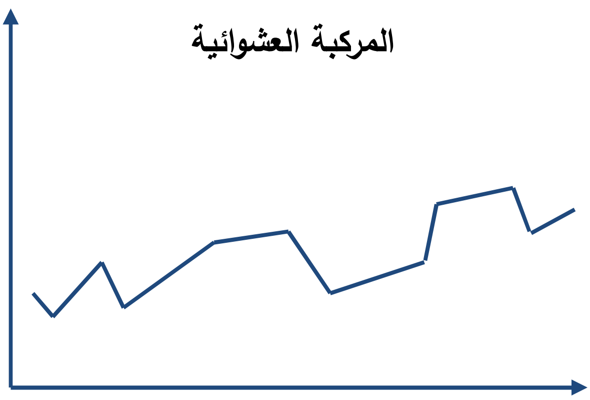
المركبة الدورية



المركبات الغير منتظمة:

أ) المركبة العشوائية (التغيرات العشوائية)

التغيرات العشوائية هي تغيرات شاذة وطارئة بمعنى أنه لا يمكن التنبؤ بوقوعها أو تحديد نطاق تأثيرها ،حيث أنها تحدث نتيجة لأسباب عارضة لم تكن في الحسبان مثل الزلازل،الفيضاناتإضراب العمال....الخ.أو هي كل التغيرات التي يمكن توقع حدوثها أو قياسها ويرمز لها بالرمز**I**، وعملية تقديرها تعتمد على امكانية تقدير بقية مكونات السلسلة **C, S, T**



3 استقرارية السلاسل الزمنية :تتصف في الغالب مختلف البيانات الاحصائية، بوجود تغيرات ممكن أن تؤثر على استقرارية السلسلة، بالتالي تحديد سكونها و تكاملها من نفس الدّرجة يعدّ أمرا مهما وتكون عملية اختبار الاستقرارية عن طريق:

- **اختبار جدر الوحدة(Unit Root Test):**

أوضحت العديد من الدراسات أنّ أغلب السلاسل الزمنية تكون مستقرة من الدّرجة الثانية، حيث إذا كانت سلسلة الفروق الأولى من سلسلة المتغير العشوائي مستقرة، نقول أنّ السلسلة مستقرة من الدّرجة الأولى، و متكاملة من الرتبة صفر، و لا تحمل جدر وحدة، و عموما فان السلسلة x_t تكون متكاملة من الدرجة(d)إذا كانت ساكنة من الدّرجة (d) لذا فانها تحتوي على عدد (d) جدر وحدةأمّا إذا كانت مستقرة بعد الحصول على سلسلة الفروق الثّانية، فنقول أنّها مستقرة من الدّرجة الثّانية.

وعلى العموم نجد أنّه لدينا عدّة اختبارات حتّى يتأكد الباحث من وجود جذر وحدة من عدمه ومنه تحديد مدى استقرارية السلسلة الزمنية، من هذه الاختبارات لدينا طريقة **ديكي فيلر المطور (Dickey-Fuller Augmentes)**، اختبار **فيليب بيرون(Philip Perrson)**،واختبار **(KPSS)**.

1) اختبار ديكي فيلر الموسع (Dickey-Fuller Augmentes)(ADF-1981):

هو أكثر تطورا من اختبار (DF) هذا الأخير الذي تمّت صياغته سنة 1979، كونه يأخذ في الحسبان عدم ترابط الأخطاء، و يضم ثلاث عناصر للتأكد من مدى استقرارية السلاسل الزمنية وهي صيغة النموذج المستخدم، حجم العينة، و مستوى المعنوية، كما أنّ يرتكز على الفرضية البديلة: $|\phi_1|<1$، وتكون صيغ تقدير النموذج باستعمال طريقة المربعات الصغرى (OLS) على النحو التالي:

النموذج (1) $$\Delta x_t = px_{t-1} - \sum_{j=2}^{p}\phi_j \Delta x_{t-j+1} + \varepsilon_t$$

النموذج (2) $$\Delta x_t = px_{t-1} - \sum_{j=2}^{p}\phi_j \Delta x_{t-j+1} + C + \varepsilon_t$$

النموذج (3) $$\Delta x_t = px_{t-1} - \sum_{j=2}^{p}\phi_j \Delta x_{t-j+1} + C + bt + \varepsilon_t$$

حيث: $P = \phi - 1$، (P درجة التأخير).

ε_t : تشويش أبيض (متوسط معدوم، وتباين θ_ε^2)

ويتم اختبار الفرضيات التالية :

$$H_0 : \phi_1 - 1 = 0$$
$$H_1 : \phi_1 - 1 < 0$$

احتسبت قيمة المقدرة (ϕ_1) من طرف **Dickey**و**Fuller**، واستخرجوا جدولاً للقيم الحرجة $(\hat{\phi}_1 - 1)$ بحيث يتم مقارنتها مع (z) المحسوبة انطلاقاً من المعادلة التالية:

$$z_{cal} = \frac{\hat{\phi}_1 - 1}{\hat{\delta}\hat{\phi}_1}$$

فإذا كانت $z_{cal} \geq z_{tab}$: معناه وجود جذر أحادي و منه نقبل الفرضية (H_0)، والسلسلة تكون غير مستقرة.

إذا كانت $z_{cal} < z_{tab}$: معناه أنّ السلسلة الزمنية مستقرة، ويمكن تحديد قيمة (P) عن طريق معيار **(Akaike)** أو معيار **(Schwarz)**، حيث:

$$Akaike(p) = n \log\left(\delta^2_{\hat{\varepsilon}t}\right) + 2(3+p)$$
$$Schawrtz(p) = n \log\left(\delta^2_{\hat{\varepsilon}t}\right) + (3+p)\log n$$

أين :

$\delta^2_{\hat{\varepsilon}t}$: تباين الأخطاء العشوائية بعد عملية التقدير.

n : عدد المشاهدات.

2) اختبار فيليبس بيرسون (Phillips Perrson) (PP -1988):

طوّر كلّ من **فيليب** و **بيرسون** اختبارات جدر الوحدة، و التي صار لها شعبية كبيرة في تحليل السلاسل الزمنية، و يختلف اختبار جدر الوحدة لـ (ADF) عن (pp) في كيفية التعامل مع ارتباط السلسلة، خاصة فيما يتعلق بالأخطاء و أخذها بعين الاعتبار، فنجد أن هذا الأخير لا يحتوي على قيم متباطئة للفروق، و يأخذ في الاعتبار الارتباط في الفروق الأولى في السلسلة الزمنية، و هذا باستخدام التصحيح المعلمي، و يسمح بوجود متوسط لا يساوي (0)، و اتجاه خطّي للزمن، ليعتبر بذلك تصحيح غير ثابت لإحصائيات

Dickey-Fuller من أجل أخذ الأخطاء بعين الاعتبار.

ويتم اختبار **Phillips** و **Perron** وفق المراحل التالية :

1.تقدر النماذج الثلاث لاختبار **Dickey-Fuller** بطريقة المربعات الصغرى العادية، وتحسب الإحصائيات المشتركة مثل (e_t) سلسلة البواقي.

2.يقدر الانحراف في المدى القصير $\hat{\delta}^2=\frac{1}{n}\sum_{t=1}^{n}e_t^2$.

3.تقدير معامل التصحيح $\left(S_t^2\right)$ (التباين في المدى الطويل)، ويحسب انطلاقاً من تباينات البواقي للنماذج المقدرة.

$$S_t^2=\frac{1}{n}\sum_{t=1}^{n}e_t^2+2\sum_{i=1}^{l}(1-\frac{i}{l+1})\frac{1}{n}\sum_{p=i+1}^{n}e_te_{t-i}$$

ولتقدير هذا التباين في المدى الطويل لابد من وجود قيمة تأخر (L) مقدرة بدلالة عدد المشاهدات:

$$n,L\approx 4\left(n/100\right)^{2/9}$$

أمّ إحصائي إحصائية PP تحسبمن خلال المعادلة التالية:

$$t_{\hat{\phi}_1}^{*}=\sqrt{k}\times\frac{\left(\hat{\phi}_1-1\right)}{\hat{\delta}_{\hat{\phi}_1}}+\frac{n\left(k-1\right)\hat{\delta}_{\hat{\phi}_1}}{\sqrt{k}}$$

مع: $k=\frac{\hat{\delta}^2}{S_t^2}$

وتكون $\left(k=1\right)$ إذا كانت سلسلة البواقي $\left(e_t\right)$ تشكل تشويش أبيض.

ويتم مقارنة إحصائية $\left(t_{\hat{\phi}_1}^{*}\right)$ مع القيمة الجدولية المستخرجة من جدول **(Mackinnon)**.

3) اختبار Kwiatkowski, Phillips, Schmidt, Shin (KPSS - 1992):

ويختلف عن الاختبارين السابقين في كون هذين الأخيرين يأخذان في الحسبان الفرضية العدمية لجدر الوحدة ضد الفرضية البديلة لجدر الوحدة، ويرتكز على استعمال مضاعف لاغرانج **(Multiplicateur de Lagrange)(LM)**بعدما يتم تقدير النموذج (2)، أو النموذج (1)، و نقوم بحساب المجموع الجزئي للبواقي حسب الصيغة التالية: $S_t=\sum_{i=1}^{t}e_i$

ثم نقوم بتقدير التباين $\left(S_t^2\right)$ في المدى الطويل بنفس الصيغة المعطاة في اختبار Phillips و **Perron** أي:

$$S_t^2=\frac{1}{n}\sum_{t=1}^{n}e_t^2+2\sum_{i=1}^{l}(1-\frac{i}{l+1})\frac{1}{n}\sum_{p=i+1}^{n}e_te_{t-i}$$

وتكون إحصائية معامل لاغرانج كما يلي:

$$LM = \frac{1}{S_t^2} \frac{\sum_{t=1}^{n} S_t^2}{n^2}$$

فإذا كانت إحصائية (**LM**)أكبر من القيمة الجدولية، نعمل على رفض الفرضية العديمة (H_0) و الذي معناه أن السلسلة الزمنية غير مستقرة.

- **اختبار التكامل المتزامن:**

و يقصد به ربط مجموعة متغيرات تكون من نفس الدّرجة، أو من درجات مختلفة، أين يؤدي هذا الربط الى الحصول على تركيبة متكاملة خطّية تكون بأقل من رتبة المتغيرات التي تمّ استعمالها في العينة أو تساويها.

كانت اسهامات كل من **Granger (1983)**، **Engele** و **Granger (1987)** موضحة لمفهوم التكامل بين متغيرين أو أكثر من النّاحية الاحصائية، خاصة في المدى الطّويل، وقد اعتمدها العديد من الاقتصاديين كمفهوم جديد ومهم في مجال الاقتصاد القياسي، وفي تحليل السلاسل الزمنية.

أ) **إختبار التكامل المتزامن باستخدام طريقة أنجل جرانجر ذات الخطوتين :**

هو من أهم الاختبارات المستعملة على السلاسل الزمنية، و يختبرها ما إذا كانت متكاملة من نفس الدرجة في مرحلتين:

المرحلة الأولى: اختبار درجة تكامل المتغيرات: من أهم شروط التكامل المتزامن أن تكون متكاملة من نفس الدرجة، و ان انعدم هذا الشرط فلن يكون لدينا تكامل، و يتم تحديد التكامل باستعمال اختبار (ADF)

و تكون السلسلتان الزمنيتان (x_{1t}) و (y_{2t}) متكاملتان إذا تحقق الشرطان التاليان:

- أن تكون كل سلسلة منتجة من سلسلة عشوائية من نفس درجة التكامل.
- أن تكون التركيبة الخطية للسلسلتين تسمح بالحصول على سلسلة من درجة تكامل أقل.

إذا كانت (x_t) سلسلة زمنية متكاملة من الدرجة (d)أي : $x_t \rightarrow I(d)$

وإذا كانت (y_t) سلسلة زمنية متكاملة من الدرجة (b)أي : $y_t \rightarrow I(b)$

فإن مجموع السلسلتين يكون متكامل بدرجة تكامل أقل أي :

$$\begin{matrix} x_t \rightarrow I(d) \\ y_t \rightarrow I(b) \end{matrix} \Rightarrow \alpha_1 x_t + \alpha_2 y_t \rightarrow I(d-b)$$

حيث: $d \geq b \geq 0$

$[\alpha_1, \alpha_2]$: عبارة عن شعاع التكامل المتزامن **(Vecteur de cointégration).**

المرحلة الثانية :اختبار Durbin-Watson (DW-1988):عبارة عن احصائية معينة تمكن الباحث من اكتشاف وجود ترابط بين الأخطاء العشوائية من عدمه، و هو اختبار من الاختبارات السهلة نسبيا يمكن استخدامه عندما يكون عدد المشاهدات صغير نسبيا، بشرط أن لا يقل عن 15 مشاهدة (n>15)،كما أنه مناسب لاختبار وجود ارتباط ذاتي من الدرجة الأولى.

وجود ارتباط بين الأخطاء العشوائية يكون له أثر كبير على مصداقية تقديرات معلمات النموذج الاحصائي، حيث لا يكون التغاير قوي، فيخلق مشكل في النموذج، و يتخذ الاختبار المعادلة التالية:

$$d = \frac{\sum_{i=2}^{n}(ei - ei - 1)^2}{\sum_{i=1}^{n} e_i^2}$$

حيث إذا كانت القيمة المحسوبة تؤول الى القيمة "**2**" فمعناه أنّ البواقي مستقرة، و منه تحقق فرضية التكامل المتزامن، أمّ إذا كانت تؤول إلى الـ"**0**" فنقر هنا بغياب التكامل.

ب) اختبار التكامل المتزامن لـ "جوهانسنجسلس"Johansen- Juselius (JJ-1988/1991)

لتحديد عدد علاقات التكامل المتزامن، نستعمل طريقة ".**Johansen**" **(1988)** المرتكزة على القيم الذاتية للمصفوفة **(les valeurs propres)** وانطلاقاً من القيم الذاتية يمكننا حساب الإحصائية λ_{trace}

$$\lambda_{trace} = -n \sum_{i=r+1}^{k} Ln(1-\lambda_i)$$

و التي يختبر من خلالها فرضية العدم التي تقول أن عدد اتجاهات التكامل المتزامن الفريدة، يقل أو يساوي العدد (q) مقابل الفرض البديل (q=r)، حيث :

n: عدد المشاهدات.

λi : القيمة الذاتية رقم (i) للمصفوفة.

k: عدد المتغيرات.

r:رتبة المصفوفة.

تتناسب العينات الصغيرة الحجم في هذا النّموذج، و يستعمل في حالة وجود أكثر من متغيّرين عكس اختبار **أنجل و جنرانجر** للتكامل المتزامن، كما يمكن أيضا من معرفة ما إذا كان هناك تكامل متزامن في حالة انحدار المتغير التابع على المتغيرات المستقلة.

(λ_{trace}) تتبع قانون احتمال يتشابه أو يقترب من توزيع (x^2) كي دو .

وقد قام كل من **Johansen** و**Juselius** بوضع جدول لقيم (λ_{trace}) سنة **(1990)**.

و بعد حساب قيمة (λ_{trace}) يتم مقارنتها مع (λ_{trace}) الجدولية)، و عليه تكون لدينا الحالات التالية:

-إذا كانت (λ_{trace}) المحسوبة $\leq$ القيمة الجدولية: نرفض الفرضية (Ho)، و نمر إلى الاختيار التالي وهكذا حتى نهاية الاختبار .

- إذا كانت (λ_{trace}) المحسوبة $\geq$ القيمة الجدولية: نقبل الفرضية (Ho)، ونتوقف عند هذا الاختبار، وتحسب عدد أشعة التكامل المتزامن على أساس الفرضية البديلة للاختبار السابق.

ج) نموذج تصحيح الخطأ Modèle à correction d'erreur (ECM):

إنّ نموذج تصحيح الخطأ يقوم على فرضية مفادها أنّ هناك علاقة توازنية في المدى الطّويل تتحدّد من خلالها القيم العشوائية، إلاّ أنّها من النّاذر أن تتحقق، و منه بالامكان أن يحصل على قيم مختلفة عن تلك التوازنية، و الفرق بين القيمتين عند كل فترة زمنية، يمثل خطأ التوازن، ليعدّل بذلك الخطأ أو جزء منه على الأقل في المدى الطّويل، لأجل هذا سمّي بنموذج تصحيح الخطأ و استخدمت معادلة التوازن التالية:

$$Y_t = \alpha + \beta X_t + \varepsilon_t$$

حيث تمثل (ε_t) 'متغير التابع وتمثلX_tمتجه من المتغيرات المفسرة. فإذا كانت مجموعة المتغيرات y_tوX_tفي حالة توازن يكون الفرق يساوي الصفر $\varepsilon_t = 0$، أين:

$.\varepsilon_t = Y_t - \alpha - \beta X_t$

و في حالة ما إذا كان الفرق غير معدوم، سيكون هناك تباعد عن التوازن، ولن تقيس القيمة ε_tالبعدعن التوازن بين X_tوY_t، ويعرف ذلك بخطأ التوازن، **Disequilibrium**.

في حالة تواجد خطأ التوازن يمكن افتراض أن Y لها علاقة مع X ومع القيم المتباطئة من X و Y ويمكن تمثيل ذلك بنموذج تصحيح الخطأ في المعادلة التالية:

$$\boldsymbol{\Delta y_{t-1} = b\Delta x_{t-1} - c(y_{t-1} - a x_{t-1}) + \varepsilon_t}$$

وتوضح هده المعادله ان التغير في y يعتمد على التغير في x وكدلك القيم المتباطئة لخطأ التوازن،بالتالي عندما تكون القيمة y_{t-1} أعلى من القيمة التوازنية فإن قيمة y_t سوف تنخفض في الفترة القادمة لتصحيح الخطأ ويعتمد ذلك على قيمة معلمة تصحيح الخطأ c بمعنى أن النموذج يقيس الكيفية التي يمكنها تصحيح القيمة للعودة إلى الوضع التوازني، أمّا a، b، فتقيسان معالم الأجل القصير والأجل الطويل، وتقيس c سرعة التكيف لتوازن الأجل الطويل.

نماذج شعاع الانحدار الذاتي(VAR)

أ) اختبار العلاقةالسببية(Causality)

يستخدم لأجل تحديد طبيعة العلاقات الموجودة بين مجموعة متغيرات في اقتصاد معيّن، وأيهما يؤثر على الآخر، كون هذه المتغيرات لا تتحرّك بنفس الاتجاه، و هذا بسبب تأثّرها بمجموعة عوامل خارجية.

قدم هذا الاختبار للسببية**Granger** سنة **1969**، و عرف العلاقة السببية بين المتغيرات في الاقتصاد على أن التغير في القيم الحالية والماضية لمتغير ما يسبب التغير في متغير آخر أي أنا لتغير في قيمx_t مثلاً الحالية والماضية يسبب التغير في قيم y_t ويتضمن اختبار جرانجر للسببية تقدير نموذج انحدار ذاتي:

$$y_t = \delta_0 + \sum_{i=1}^{p} \delta_i\, y_{t-i} + \sum_{j=0}^{q} \lambda\, x_{t-j} + u_t$$

$$x_t = \alpha_0 + \sum_{i=1}^{n} \alpha_i\, x_{t-i} + \sum_{j=0}^{m} \beta\, y_{t-j} + v_t$$

حيث أن$(\lambda\ ,\delta i\ ,\beta\ ,\ \alpha i)$ معلمات يراد تقديرها،U_t وV_t حدين عشوائيين بتباين ثابت ومتوسط حسابي يساوي الصفر ويتم تقدير المعادلتين باستخدام طريقة المربعات الصغرى **(OLS)**.

و لاختبار العلاقات السببية سوف نستعمل الفرضيتين العديمتين التاليتين:

$$\begin{Bmatrix} H_0: \lambda = 0 \\ H_0: \beta = 0 \end{Bmatrix}$$

إن لم نستطيع رفض أي من هاتين الفرضيتين، فإن المتغيرين المدروسين Xو Y مستقلين عن بعضهما، أما إذا تم رفضهما معاً فهناك علاقة سببية في الاتجاهين (Xيسبب Y و Y يسبب X).

نستخدم إحصائية فيشر المحسوبة و الجدولية لاختبار الفرضيتين، فإذا كانت F^* أكبر من إحصائية فيشر F الجدولية عند مستوى معين من المعنوية، نرفض الفرضية العديمة، و معناه وجود علاقة سببية باتجاهين أ ي أنّ المتغير x يتأثر بالمتغير y، و العكس صحيح، أمّا إذا كانت أصغر فنقبل الفرضية العديمة أي عدم وجود علاقة سببية بين X و Y.

و تقدّر إحصائية فيشر المحسوبة F^* وفق القانون التالي:

$$F^* = \frac{(SCRR - SCRU)/C}{SCRU(n-k-1)}$$

حيث أن: $SCRR$: هو مجموع بواقي المربعات في المعادلة المختزلة.

$SCRU$: هو مجموع بواقي المربعات في المعادلة غير المختزلة.

C: هو عدد المعاملات المختزلة.

k: هو عدد المتغيرات الأصلية (بدون اختزال) في المعادلة.

n: هو عدد المشاهدات المستخدمة لتقدير المعادلة الغير مختزلة.

إنّه و باستعمال معامل الإرتباط (R)يمكن للباحث أن يحدّد طبيعة العلاقة بين المتغيرات (Xو Y) المستقل و التابع، و إذا ما كانت العلاقة بينهما ضعيفة، معناه التغاير ضئيل، و سيكون هناك تغاير بين المتغيرات بوجود ارتباط، لكن هذا الأخير ليس معناه السببية، لنستنتج أنّ هذا الاختبار تمّ استخدامه لمعرفة أو تحديد نوع و اتجاه العلاقة بين المتغيرات الاقتصادية.

ب) تحليل دوال الاستجابة الدفعية:

تستعمل هنا نماذج **(VAR)**في تحليل الاستجابة الدفعية، و تقيس هذه الدّوال أثر الاستجابة للقيم الحالية و المستقبلية للعوامل الدّاخلة في النّموذج، أين يعتمد هنا على تأثير صدمة البواقي (ε_t) على القيم الحالية و المستقبلية للمتغيرات التّابعة، وذلك بمقدار انحراف معياري واحد للصّدمة في مكون النّموذج، و هذا ما يساعد و يمكن الباحث من معرفة تأثير تغير العامل الاقتصادي على نفسه، و على باقي متغيرات النّموذج، و هذا خلال فترة زمنية معيّنة و في المستقبل أيضا، و سينتقل هذا التأثير إلى المتغيرات الأخرى عن طريق هيكل ديناميكية نماذج (VAR) ، و إذا ما افترضنا النموذج التالي:

$$X_t = \alpha_x + \sum_{i=1}^{p} \beta_{x,i} X_{t-i} + \sum_{i=1}^{p} \emptyset_{x,i} Y_{t-i} + \varepsilon_{x,t}$$

$$Y_t = \alpha_y + \sum_{i=1}^{p} \beta_{y,i} Y_{t-i} + \sum_{i=1}^{p} \emptyset_{x,i} X_{t-i} + \varepsilon_{y,t}$$

حدوث صدمة في حد الخطأ ε_x سيؤثر حتماً في القيمة الحالية لـX_t كما انه سيؤثر في القيم المستقبلية لكل من **X**و **Y** نظراً لاحتواء المعادلتين على القيم السابقة لـ **X**. فإذا افترضنا أن هذه الصدمة في ε_x مقدرة بـ 1% فإن سينتج عن ذلك التأثير كما في الجدول التالي:

أثر صدمة في البواقي على القيم الحالية والمستقبلية للمتغيرات التابعة.

الفترة	ديناميكية تأثير الصدمة
في الفترة t:	$\begin{bmatrix} \Delta X_t \\ \Delta Y_t \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$
في الفترة t+1:	$\begin{bmatrix} \Delta X_{t+1} \\ \Delta Y_{t+1} \end{bmatrix} = \begin{bmatrix} \beta_{x,i} \cdots & \emptyset_{x,i} \\ \beta_{y,i} \cdots & \emptyset_{x,i} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$
في الفترة t+2:	$\begin{bmatrix} \Delta X_{t+2} \\ \Delta Y_{t+2} \end{bmatrix} = \begin{bmatrix} \beta_{x,i} \cdots & \emptyset_{x,i} \\ \beta_{y,i} \cdots & \emptyset_{x,i} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} c \\ d \end{bmatrix}$
....	
....	....

التحليل العاملي : Analysis Of Factor

1) تعريف التّحليل العاملي: هو أسلوب العلاقات المتبادلة بين عدد من المتغيّرات. تمّ وضعه من قبل "ج. ب. بنزكري"(**J.P.Benzecri**)سنة **1990** ، يهدف إلى تقليص عدد المتغيّرات في شكل تركيبات أساسية يسهل استعابها وترجمتها، ويعتبر هذا الأسلوب هو المناسب للدّراسات المتعدّدة المتغيّرات.

***التّحليل العاملي التّأكيدي(-Confirmatory Factor Analysis-CFA):**هو شكل خاص من أشكال التّحليل العاملي، يستخدم لاختبار ما إذا كانت المتغيّرات تناسب النّموذج المفاهيميالإفتراضي، ويستند على النظريات والبحوث التّجريبية السّابقة، وعادة ما تستخدم نمذجة المعادلات الهيكلية(**SEM**) لتحقيق التّحليل العاملي التّأكيدي، هذا الأخير الذي يستخدم كخطوة أولى لتقييم نموذج القياس المقترح في نموذج المعادلات الهيكلية من خلال مجموعة من المؤشّرات.

***التّحليل العاملي الاستكشافي(-Exporatory Factor Analysis-EFA):** ويسمح بتحديد مجموعة المتغيّرات(**Items**) التي تختلف فيما بينها، وتظهر لتمثيل المتغيّرات الكاملة، بعبارة أخرى فإنّ "**EFA**" يمكّن من استكشاف العلاقة بين المتغيّرات المقاسة لتحديد ما إذا كانت هذه العلاقة يمكن تلخيصها بعدد قليل ومهم من التّركيبات الكامنة.

2)استخدام التّحليل العاملي:

يتم استخدام التّحليل العاملي لتحديد عناصر التركيبة وتحديد كيف يمكن لكل منها(تركيبة) أن تفسر كلّ عنصر، ويمكّن لهذا التّحليل منتحقيق هدفين أساسيّين:

- تلخيص البيانات: التحليل العاملي يكشف عن الأبعاد الكامن التي يتم ترجمتها مرّة واحدة، ويصف البيانات في شكل ملخّص.
- تقليص البيانات: يحسب عدد النّقاط بالنّسبة لكل بعد ويحل محلّ المتغيّرات الأصلية.

في حين أنّ الأساليب الأخرى(الانحذار، تحليل التّباين،...إلخ) المتغيرات بها تعتبر إمّا مستقلّة أو تابعة، كما نجد أنّ في التحليل العاملي جميع المتغيّرات مرتبطة ببعضها البعض، وتشكّل تركيبات تضم عدد من المتغيّرات وهذا من أجل تعظيم تفسير جميع المتغيّرات.

3)حجم العيّنة اللاّزم للتحليل العاملي:

في هذه الطّريقة يجب ألاّ يقل عدد المشاهدات عن الخمسين**(50)** ومن الأفضل ألاّ يقل عن مئة**(100)** فرد.

4)مؤشّرات قياس التحليل العاملي:

قبل القيام بالتّحليل العاملي ، من المهم التأكّد من أن البيانات مقبولة، كما يجب أن تشكل المتغيّرات مجموعة متجانسة من أجل الحصول على تركيبات مشتركة، وأن تكون مصفوفة البيانات مرتبطة بما فيه الكفاية لتحقيق نتائج**(EFA)** المرغوب فيها، ويوجد عدّة مؤشّرات لاثبات ذلك:

***اختبار ملائمة اختيار العيّنة –Kaiser Meyer OlkinOf Sampling Adequacy-KMO**

ويوفّر هذا المؤشّر نظرة عامّة عن نوعية العلاقات المتبادلة بين فقرات الاستبيان**(Items)** ومدى تماسك وتجانس المتغيّرات، ويختبر إذا ما كان الإرتباط الجزئي بين المتغيّرات ضعيف.

تتراوح قيم المؤشّر **"KMO"** بين الـ **"0"** و **"1"** وتفسيره على النّحو التّالي:

أقل من 0,50	**غير مقبول**
أكثر من 0,50	**غير كــافي**
أكثر من 0,60	**مقبول**
أكثر من 0,70	**جيّد**
أكثر من 0,80	**ممتاز**

***اختبار بارتليت(Bartlett Test Of Sphericity):** اختبار احصائي يسمح بقياس مالإذا كانت مصفوفة الارتباط كاملة معرّفة، أي وجود ارتباط بين جميع عناصرها، وتؤكّد ذلك مستوى المعنوية(P<0,05) من أجل رفض الفرضية العدمية.

***اختبار الاشتراكية(Communality Test):** ويقيس نسبة التباين المفسرللعبارت(Items)، و حسب **Carricano et Poujol** يجب أن تتجاوز قيمتها الـ **0,50** ومن الأفضل أن تتعدّى الـ **0,70**.

وحسب الباحث "هير"(Hair, 2010) فإنّجميع المعاملات التي تقل قيمتها عن الـ **0,50** يجب حدفها لأنّ ليس لها أي تفسير و **يمكن تحديد قيم اختبار الاشتراكية في الجدول التالي:**

أقل من 0,50	غير مقبولة
0,60–0,70	جيّدة
أكثر من 0,80	ممتازة

5) استخراج عدد التّركيبات:

يوجد عدّة اجراءات لتحديد العدد الأمثل من التّركيبات التي تضم جميع المتغيّرات، ولكن غالبا ما تعتمد نسبة التّباين لكل عبارة(Items) أهمّها:

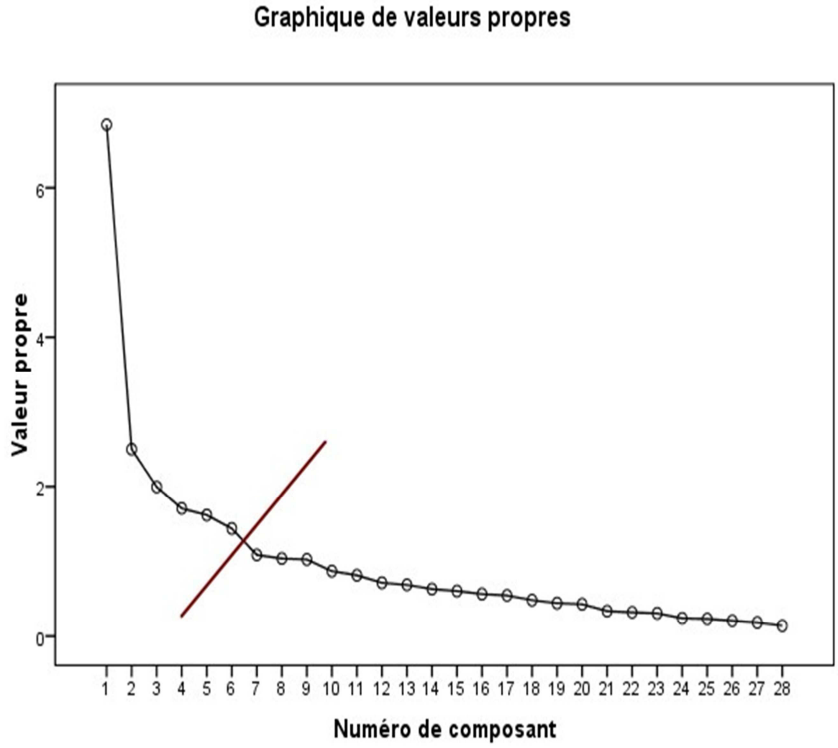
***اختبار القيم الذّاتية(Kaiser's Eigenvalue Than One Rule):** وهو المعيار الأكثر استخداما على نطاق واسع.

القيم الذّاتية التي تفوق الواحد الصّحيح(> 1) تفسّر جزءا كبيرا من التّباين الكلّي، وتمثّل مكوّنات النّموذج.

***اختبارCattell's(Cattel's Scree Test):** أحد الاختبارات المهمّة لاختيار عدد المكوّنات هو اختبار "**Cattel, 1966**"، والذي يعتمد على تمثيل بياني للقيم الذّاتية في شكل نقاط مرتبة ترتيبا

تنازليا، مرتبطة ببعضها البعض بواسطة خطّ، ويمثّل كل نقطة قيم العبارات(**Items**) بحيث يتم الاحتفاظ بالقيم التي تقع قبل تغير المنحنى، وتمثّل القيم المقبولة أو الممثّلة للنّموذج وتلغى تلك التي تليها.

منحنى**Cattel**



6) تحديد مكوّنات التّركيبة(Ratation Of Factors):

من أجل تحديد مكوّنات التّركيبات، لابد من إجراء عملية التدوير"**Ratation Of Factors** "، الأمر يسمح بتحديد مجموعة من المتغيّرات المرتبطة ببعضها البعض، كما يضمن لنا انتماء كل عبارة (**Items**) الى تركيبة واحة فقط.

يتحقّق ذلك من خلال إعادة توزيع تباين التّركيبات المستخرجة من أجل تحقيق"تركيبةعاملية بسيطة"(**Factor Structure**)، فعندما تثبت المحاور على زاوية تسعين درجة (**90°**)، تسمّى بـ"الدّوران المتعامد" (**Orthogondl Factor Rotation**)، أمّا إذا كانت المحاور غير مقيّدة، تسمّى بالدّوران المائل (**Oblique Factor Rotation**).

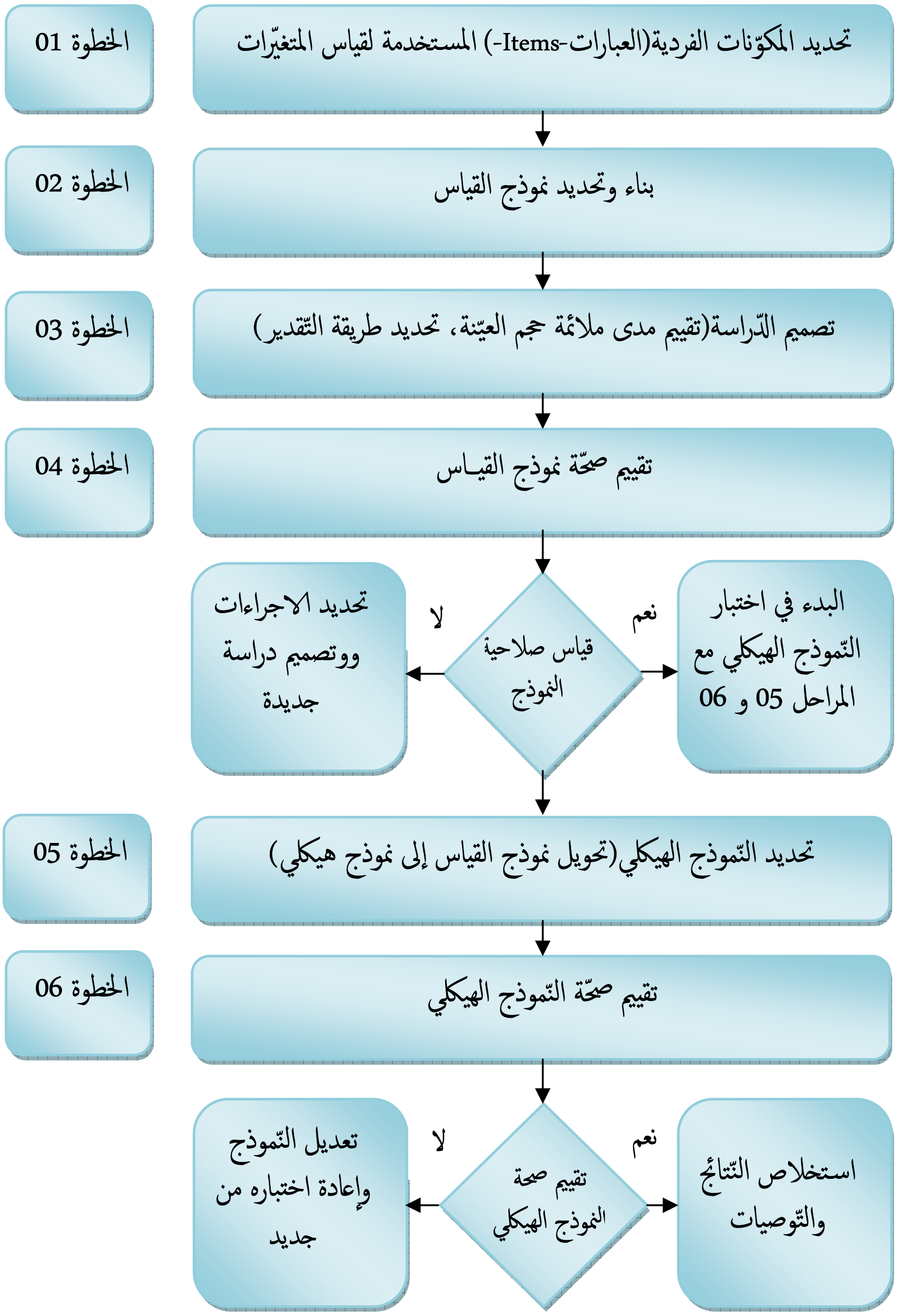
في الممارسة العملية يفضّل استخدام وبانتظام طريقة الدّوران المتعامد (**Orthogondl Factor Rotation**)إذا ما كان الهدف منه تقليص عدد المتغيّرات في مصفوفة البيانات.

التّحليل باستخدام المعادلات الهيكلية:(-Structural Equation Modeling-SEM):

1) **التّعريف:**عبارة عن أحد النّماذج الاحصائية التي تسعى إلى اختيار العلاقات المتبادلة بين المتغيّرات المتعدّدة، وتبرز هيكل العلاقات في شكل معاملات، ومن أهم ما يميّز هذا النّوع من النّماذج هو معالجة العلاقات المتعدّدة في آن واحد، كما يعتبر من النّماذج الاحصائية القوية التي تمكّن من التّعامل مع النّماذج المعقّدة.

2) **خطوات بناء نموذج المعادلات الهيكلية:** حددت خطوات بناء نموذج المعادلات الهيكلية في ستة خطوات اساسية:

- ✓ وضع نموذج مفاهيمي يستند على مراجعة الأدبيات.
- ✓ بناء تمثيل بياني للعلاقات السّببيّة.
- ✓ تحويل التّمثيل البياني إلى مجموعة من المعادلات الهيكلية.
- ✓ اختيار نوع الإدخال المصفوفي وتقدير النّموذج.
- ✓ تقييم النّموذج المقترح.
- ✓ تقييم نتائج الإختبار(**Goodness Of Fit**)وإجراء تعديلات على النّموذج.



خطوات نمذجة المعادلات الهيكلية

3. مؤشرات اختبار صلاحية النموذج:

يمكن تلخيص المؤشرات الاكثر استخداما لاختبار صلاحية النّموذج في الجدول التّالي:

المؤشّرات	الرمز	نوع القياس	القيم المعيارية
Chi Square	χ^2	Model Fit	χ^2, **df**, P$\geq 0,05$
Normed Chi Square	χ^2/df	Absolute Fit And Parsimony Of Model	$1,0 < \chi^2$/**df**$< 3,0$
Goodness Of Fit Index	GFI	Absolute Fit	Volue$\geq 0,90$
Adjusted Goodness Of Fit Index	AGFI	Absolute Fit	Volue$\geq 0,90$
Standarised Root Mean Square Residnal	RMR	Absolute Fit	Volue$\leq 0,10$
Root Mean Square Error Approximation	RMSEA	Absolute Fit	Volue$\leq 0,10$
Incrementd Index Of Fit	IFI	Incremental Fit	Volue$\geq 0,90$
Tucker Lewis Index	TLI	Incremental Fit	Volue$\geq 0,90$
Comparative Fit Index	CFI	Incremental Fit	Volue$\geq 0,90$
	AIC	Parsimony Fit	Lowest Possible
	CAIC	Parsimony Fit	Lowest Passible

المراجع باللغة العربية :

1. أبو الصالح محمد صبحي و أمجد ضيف الله الناصر، 2011، دليل التحليل الإحصائي بالستخدام spss، دار اليازوري العلمية للنشر و التوزيع ،الأردن .
2. أبو صالح محمد صبحي وعوض عدنان محمد، 2004، مقدمة في الإحصاء، دار المسيرة للنشر والتوزيع، الأردن.
3. أبو صالح محمد صبحي، 2000، الطرق الإحصائية، دار اليازوري العلمية للنشر والتوزيع الأردن.
4. تومي صالح، 2011، مدخل لنظرية القياس الاقتصادي دراسة نظرية مدعمة بأمثلة وتمارين الجزء الأول، ديوان المطبوعات الجامعية.
5. حسين علي بخيت و سحر فتح الله، 2007، الاقتصاد القياسي، دار اليازوري العلمية للنشر والتوزيع، الأردن.
6. حسين علي بخيت و غالب عوض الرفاعي،2007، تحليل ونمذجة البيانات باستخدام الحاسوب- تطبيق شامل للحزمة SPSS، الأهلية للنشر والتوزيع، عمان، الأردن.
7. حمزة محمد دودين، 2010، التحليل الاحصائي المتقدم للبيانات باستخدام spss، دار المسيرة للنشر و التوزيع و الطباعة، الأردن.
8. شيخي محمد، 2012، طرق الاقتصاد القياسي محاضرات وتطبيقات، دار الحامد للنشر والتوزيع، عمان، الأردن.
9. عبد الحميد عبد المجيد البلداوي،2007، أساليب البحث العلمي و التحليل الإحصائي، دار الشروق للنشر و التوزيع، الأردن.

10.عبد القادر محمد عبد القادر عطية، 2000، الاقتصاد القياسي بين النظرية والتطبيق، الدار الجامعية، مصر.

11.عبد القادر محمد عبد القادر، 2005، الحديث في الاقتصاد القياسي بين النظرية والتطبيق الدار الجامعية للنشر والتوزيع، مصر.

12.عز حسن عبد الفتاح، 2013، مقدمة في الإحصاء الوصفي والاستدلالي باستخدام SPSS خوارزم العلمية، العربية السعودية، جدة.

13.محفوظ جودة، 2008، التحليل الاحصائي باستخدام spss، دار وائل للنشر، الأردن.

14.محمد خير سليم أبو زيد، 2010، التحليل الاحصائي للبيانات باستخدام spss، دار الصفاء للنشر و التوزيع، الاردن.

15.محمد عبد العال النعيمي وحسن ياسين طعمة، 2008، الاحصاء التطبيقي ، دار وائل للنشر

16.محمود محمد سليم ، 2008، مقدمة في الإحصاء، مكتبة المجتمع العربي، الأردن.

17.مكيد علي ،2007،الاقتصاد القياسي دروس ومسائل محلولة، ديوان المطبوعات الجامعية الجزائر . بن عكنون. والتوزيع، الاردن.

المرجع باللغة اجنبية :

1. Morgenthaler. Stéphane, 2001, Introduction à la statistique, Edition Presses polytechniques et universitaires, Suisse, 2 ème édition.
2. Pupion. Pierre Charles, 2004, Statistique pour la gestion, Edition Dunod, France.
3. Lethielleux. Maurice, 2005, Statistique descriptive, Edition Dunod, France, 4 ème édition.
4. Khedhiri. Sami, 2005, Cours d'introduction à l'économétrie, Centre des publications universitaires, Tunisie.
5. Cadoret. Isabelle; Benjamin. Catherine; Martin. Franck; Herrard.
6. Nadine

   et Tanguy. Steven, 2004, Econométrie appliquée, Edition De Boeck, Belgique
7. Bourbonnais. Régis, 2004, Econométrie, Edition Dunod, France, 5ème édition.
8. Bourbonnais REGIE, 2002, Econométrie : Manuel Et Exercices Corrigés 4e Edition, Dunod Paris.
9. Tribout. Brigitte, 2007, Statistique pour économistes et gestionnaires Pearson éducation, France.
10. Hair Jr Joseph F, Black William C, Barry J. Babin, Anderson Rolph E, 2010, Muti Variate Data Analyse, Pearson Prentice Hall, Seventh Edition.
11. Jakobowicz Emmanuel, 2006, Les modèles d'équations structurelles à variables latenteslatentes, Cours de Statistique Multivariée Approfondie Conservatoire National Des arts et Métiers, France.
12. Maru Carricano , Fanny poyal , 2008, alnalyse de données avec spss, collection synthex.