الجمهورية الجزائرية الديمقراطية الشعبية
المركز الجامعي عبد الحفيظ بوالصوف ميلة
معهد العلوم الاقتصادية والتجارية وعلوم التسيير
قسم: علوم التسيير

رقم المطبوعة:........../2025

مطبوعة بيداغوجية بعنوان :

النمذجة الإحصائية

موجهة لطلبة السنة الأولى، طور الماستر، شعبة: علوم التسيير، تخصص: إدارة أعمال

من إعداد الدكتور: بوبكر ياسين
الرتبة: أستاذ محاضر قسم أ

السنة الجامعية 2025/2024

مقدمة

مقدمـــة

في ظل التطورات المتسارعة في العلوم والبحث الأكاديمي، أصبح استخدام الأساليب الإحصائية الحديثة ضرورة لا غنى عنها لتحليل البيانات واتخاذ القرارات المستندة إلى الأدلة. ومن هذا المنطلق، تهدف هذه المطبوعة البيداغوجية إلى تزويد الطلبة بمفاهيم أساسية ومتقدمة في مقياس النمذجة الإحصائية، مع التركيز على الجوانب التطبيقية التي تعزز من قدرتهم على التعامل مع البيانات بطرق منهجية وفعالة.

وتتمحور هذه المطبوعة حول تقديم مبادئ النمذجة الإحصائية، بما في ذلك أسس بناء النماذج، اختيار المتغيرات، اختبار الفرضيات، وتحليل النتائج. كما تسعى إلى تبسيط المفاهيم الإحصائية من خلال أمثلة عملية وتمارين تطبيقية تُمكّن الطلبة من فهم كيفية توظيف النماذج الإحصائية في مختلف التخصصات التابعة لميدان العلوم الاقتصادية.

كما تأخذ هذه المطبوعة بعين الاعتبار تحديات تعلم النمذجة الإحصائية، حيث تم إعداد محتواها بناء على المحاور المدونة في المادة التعليمية بأسلوب واضح ومنهجي، متدرج من المفاهيم الأساسية إلى التقنيات الأكثر تقدماً، مما يسمح للطالب باستيعاب المعرفة بطريقة سلسة ومتسلسلة. كما نأمل في تزويد الطلبة بالأدوات والمهارات التي تمكنهم من تحليل البيانات بكفاءة، وتفسير النتائج بدقة، واتخاذ قرارات مبنية على أسس علمية، مما يعزز من قدراتهم البحثية في اعداد مذكراتهم والمهنية في المستقبل.

أما عن محتويات المطبوعة، فلقد تم تقسيمها حسب البرنامج المقرر إلى ثمانية محاور أساسية وهي:

المحور الأول: مقدمة في النمذجة الإحصائية (مفهوم النموذج، أنواع النموذج، تخصيص النموذج)

المحور الثاني: تحليل الانحدار الخطي البسيط (تحديد قيم معلمات النموذج؛ اختبار الموثوقية، التنبؤ)

المحور الثالث: تحليل الانحدار الخطي المتعدِّد (خطوات صياغة نموذج متعدِّد، تقدير معلمات النموذج، دراسة صلاحية النموذج)

المحور الرابع: الارتباط الجزئي، الازدواج الخطي وطرق اختيار المتغيرات التفسيرية

المحور الخامس: المشاكل القياسية: الارتباط الذاتي للأخطاء، عدم ثبات تباين الأخطاء، التوزيع غير الطبيعي للأخطاء

المحور السادس: عموميات حول السلاسل الزمنية وتقدير مركباتها

المحور السابع: الاستجرارية والارتباط الذاتي والجزئي

المحور الثامن: نموذج تمهيد الأسي للتنبؤ بالسلاسل الزمنية.

المحور الأول:

مقدمة في النمذجة الإحصائية (مفهوم النموذج، أنواع النموذج، تخصيص النموذج)

1- تعريف الاقتصاد القياسي، Definition of Econometrics:

يعد الاقتصاد القياسي Econometrics أسلوب من أساليب التحليل الاقتصادي يهتم بالتقدير العددي (الكمي) للعلاقات بين المتغيرات الاقتصادية معتمدًا في ذلك على النظرية الاقتصادية Economic Theory، والرياضيات Mathematics، والإحصاء Statistics، للوصول إلى هدفه الخاص باختبار الفروض والتقدير ومن ثم التنبؤ بالظواهر الاقتصادية.

هذا يعني أن الاقتصاد القياسي يحاول الاستعانة أولاً بالنظرية الاقتصادية لتحديد المشكلة المراد دراستها ولأهم المتغيرات الاقتصادية والاجتماعية التي تؤثر فيها، ومن ثم يستعين بالاقتصاد الرياضي لتوصيف العلاقات القائمة بين المتغيرات في شكل رموز ومعادلات، وأخيرًا يستعين بعلم الإحصاء فيستفيد منه في تطوير واستنباط طرق القياس لتقدير معالم الصّيغ المقترحة واختبار الفروض ومن ثم الوصول إلى النتائج الدقيقة التي يمكن الاعتماد عليها في التنبؤ بالمشكلة المدروسة.

بذلك يمكن القول بأنَّ الاقتصاد القياسي هو تكامل للنظرية الاقتصادية مع الرياضيات والأساليب الإحصائية بهدف اختبار الفروض عن الظواهر الاقتصادية، وتقدير معاملات العلاقات الاقتصادية والتنبؤ بالقيم المستقبلية للظواهر الاقتصادية.

عليه يمكن تعريف الاقتصاد القياسي بأنه علم اجتماعي تستخدم فيه أدوات النظرية الاقتصادية والرياضيات والإحصاء لتحليل الظواهر الاقتصادية، وأنه يتكون من كلمتين أصلهما إغريقي Economy اقتصاد و Metrics والتي تعني قياسات. ويعرف البعض الاقتصاد القياسي على أنه فرع المعرفة الذي يهام بقياس العلاقات الاقتصادية من خلال بيانات واقعية بغرض اختبار مدى صحة هذه العلاقات كما تقدمها النظرية، أو تفسير بعض الظواهر أو رسم بعض السياسات أو التنبؤ بسلوك بعض المتغيرات الاقتصادية (عبد القادر عطية، 2009 صفحة 4).

2- أهداف الاقتصاد القياسي The Goals of Econometrics:

يمكن التعرف على ثلاث أهداف أساسية للاقتصاد القياسي هي:

1.2 تحليل واختبار النظريات الاقتصادية المختلفة:

إن تحليل واختبار النظريات الاقتصادية، يعد هدفًا رئيسًا من أهداف الاقتصاد القياسي، ولا يمكن عد النظرية الاقتصادية صحيحة ومقبولة ما لم تجتز اختبارًا كميًا عدديًا يوضح قوة النموذج ويفسر قوة العلاقة بين المتغيرات الاقتصادية.

2.2 رسم السياسات واتخاذ القرارات

يساهم الاقتصاد القياسي برسم السياسات واتخاذ القرارات عن طريق الحصول على قيم عددية لمعلمات العلاقات الاقتصادية بين المتغيرات لتساعد رجال الأعمال والحكومات في اتخاذ القرارات الحالية من حيث توفيره لصيغ وأساليب مختلفة لتقدير المرونات والمعلمات الفنية والتكلفة الحدية والإيرادات الحدية، والميل الحدي للاستهلاك والادخار والاستثمار وغير ذلك. وتأسيسًا على ذلك فإن معرفة القيم العددية لمعلمات النموذج المقدر تساعد على إجراء المقارنات واتخاذ القرار المناسب سواءً على مستوى المنشأة أو الدولة.

3.2 التنبؤات بقيم المتغيرات الاقتصادية في المستقبل:

يساعد الاقتصاد القياسي رجال الأعمال والحكومات في وضع السياسات من خلال توفير القيم العددية لمعلمات Parameters، المتغيرات الاقتصادية والتنبؤ بما ستكون عليه الظاهرة الاقتصادية مستقبلاً. إن هذه التنبؤات تمكن واضعي السياسة ومتخذي القرار لتنظيم الحياة الاقتصادية واتخاذ إجراءات معينة للتأثير في متغيرات اقتصادية معينة، مثال ذلك، لو أرادت الحكومة معرفة الآثار المحتملة للسياسة النقدية على التضخم والبطالة، وما هو الأثر المتوقع لزيادة أسعار السلع البديلة أو المكملة على الكمية المطلوبة من السلعة الأصلية، حيث أن الاقتصاد القياسي سوف يحدّد مستوى الكمية فيما إذا كان مرتفعًا أو منخفضًا وهكذا لبقية الظواهر الاقتصادية وما يتعلق بها مستقبلاً.

3- علاقة الاقتصاد القياسي بالعلوم الأخرى:

للاقتصاد القياسي علاقة وثيقة بالنظرية الاقتصادية والاقتصاد الرياضي، الإحصاء الاقتصادي، والإحصاء الرياضي، إن هذه الفروع تتكامل من أجل توفير قيم عددية لمعلمات المتغيرات الاقتصادية المختلفة، إلا إن أياً من هذه الفروع لا يعد بديلاً عن الاقتصاد القياسي، ويمكن توضيح ذلك كما يلي:

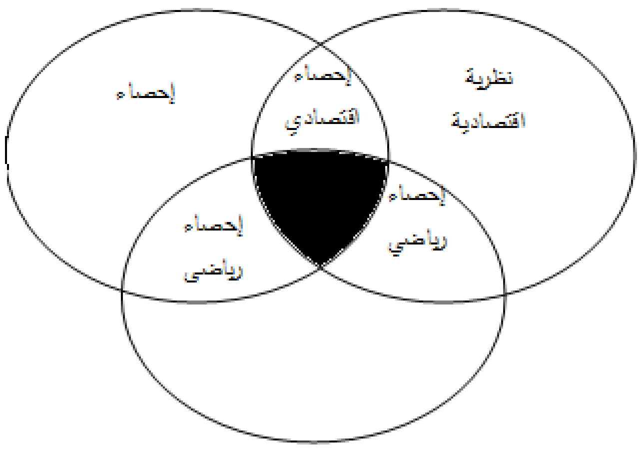
1- تقوم النظرية الاقتصادية بدراسة العلاقة بين المتغيرات الاقتصادية فتنص النظرية الاقتصادية الجزئية مثلاً على أن زيادة سعر سلعة ما يسبب انخفاضًا في الطلب عليها، فتفترض هذه النظرية وجود علاقة عكسية بين السعر والكمية المطلوبة من السلعة، ولكنها لم تعط أي قياس عددي للعلاقة بين هذين المتغيرين، فلم تبين مقدار الانخفاض للكمية المطلوبة المصاحب لتغير معين في السعر. فتصبح هذه المهمة من مهمات الاقتصاد القياسي بعد توصيفه رياضيًا. بذلك يمكن القول أن العلاقة بين المتغيرات الاقتصادية المستوحاة من النظرية الاقتصادية تبقى مسألة مجردة ما لم يتم تقديرها أي تقدير معالمها في ضوء البيانات الإحصائية الواقعية والتي هي من مهمات القياس الاقتصادي (تحديد الطابع الكمي للعلاقات بين المتغيرات الاقتصادية الجارية في واقع معين وذلك بالاسترشاد بالنظرية الاقتصادية).

2- يهتم الاقتصاد الرياضي بإعادة صياغة العلاقة التي تم تحديدها بالاعتماد على النظرية الاقتصادية رياضيًا أي على هيئة معادلات ورموز رياضية بدون قياس أو برهنة عددية لتلك الصياغات، فالقياسات والبرهنة العددية هي من مهمات القياس الاقتصادي.

3- الإحصاء الاقتصادي يقتصر دوره على تجميع البيانات الإحصائية الخاصة بالمتغيرات الاقتصادية التي تتكون منها العلاقات المحدَّدة في (1) و (2) أعلاه وتسجيلها وجدولتها أو رسمها، وينصبُّ دور القياس الاقتصادي على تحليل واختبار نوع العلاقة بين المتغيرات بهدف معرفة مدى مطابقة النتائج مع منطوق النظرية الاقتصادية.

4- أما مادة الإحصاء الرياضي فهي تجهز الباحث بأدوات تحليلية يستخدمها في دراسة العلاقات بين المتغيرات الاقتصادية وبطرق خاصة لمعالجة أخطاء التقدير تمهيدًا لاستخدامها في تحقيق أهداف القياس الاقتصادي.

لذلك يمكن النظر إلى علم القياس الاقتصادي على أنه نقطة التقاء ثلاث علوم رئيسة هي: الاقتصاد والرياضيات والإحصاء كما مبين في الشكل (1.1).



الشكل (1.1): الاقتصاد القياسي وعلاقته بالعلوم الأخرى.

يتضح ممَّا تقدم أن النظرية الاقتصادية تعطينا فكرة عامة حول شكل العلاقة بين المتغيرات الاقتصادية ويأتي دور الاقتصاد القياسي لتحديد المقدار الكمي لتلك العلاقة بالاعتماد على الاقتصاد الرياضي الذي يحاول تصوير العلاقة المذكورة بشكل معادلة رياضية، وطرق الإحصاء الرياضي لملاءمتها لطبيعة العلاقة القائمة، كل ذلك يتحقق بالاعتماد على الإحصاء الاقتصادي الذي يغذي القياس الاقتصادي بالمادة الأولية اللازمة للتحليل في صورة بيانات مجمعة ومبوبة وبدون هذه البيانات تصبح عملية القياس أمرًا مستحيلاً.

4- النموذج الاقتصادي، Economic Model:

يعرف النموذج الاقتصادي بأنه مجموعة من العلاقات بين المتغيرات الاقتصادية لتمثيل ظاهرة معينة بصورة خالية من التفاصيل والتعقيدات ولكنها ممثلة للواقع بهدف تعليلها أو التنبؤ بها والسيطرة عليها. وقد يتكون من معادلة واحدة، Single Equation مثل معادلة الطلب أو معادلة العرض ويسمى عندئذٍ النموذج بكونه نموذج ذا معادلة منفردة، أو من مجموعة من المعادلات وتسمى بالمعادلات الآنية، Simultaneous Equation، كنموذج السوق.

وقد يكون الهدف من النموذج هو تقدير قيم عددية لمعلمات علاقة بين متغيرات اقتصادية بغية التنبؤ أو تحليل هيكل اقتصادي أو تقييم سياسة اقتصادية. ويستخدم النموذج الاقتصادي الرموز الرياضية، فمثلاً تفترض النظرية الاقتصادية بأن الاستهلاك C، دالة في الدخل، Y، أي أن:

$$C = f(Y) \ldots \ldots \ldots \ldots (1.1)$$

إذ تمثل C متغير الاستجابة Dependent variable، أما Y فيمثل المتغير التوضيحي Independent variable، وتحكم العلاقة بين المتغيرات الاقتصادية بعدد من الصيغ أبسطها الصيغة الخطية، فبتحويل العلاقة (1.1) إلى صيغتها الخطية تكون:

$$C = B_0 + B_1Y \ldots \ldots \ldots \ldots \ldots (2.1)$$

حيث تمثل B_0، B_1 المعلمات Coefficients، ويمكن أن توضح من الناحية الرياضية على النحو الآتي:

B_0: تمثل معاملة التقاطع وهي عبارة عن المسافة العمودية المحصورة بين نقطة الأصل ونقطة تقاطع خط الانحدار مع المحور العمودي، وتمثل B_0 قيمة C عندما تكون قيمة Y مساوية للصفر ويطلق عليها بالاستهلاك الذاتي.

B_1: تمثل ظل الزاوية التي يصنعها خط الانحدار مع مستوى الأفق وتسمى بالميل، Slope، الحدي لخط الانحدار، وتمثل قيمة B_1 الزيادة الحاصلة في قيمة المتغير التابع، C، نتيجة زيادة المتغير التوضيحي Y، بمقدار وحدة واحدة. بعبارة أخرى إذا زاد الدخل Y بمقدار وحدة واحدة من Y_1 إلى Y_2 يصبح قيمته $(Y_1 + 1)$ يترتب على ذلك زيادة في قيمة الاستهلاك، C، من C_1 إلى C_2 . ولنفترض بمقدار ΔC. وبالإمكان إثبات أن الزيادة في المتغير التابع ΔC يساوي قيمة الميل B_1، رياضيًا وبيانيًا

أ-رياضيـًا: $C = B_0 + B_1Y$

عندما يزداد المتغير المستقل Y بمقدار وحدة واحدة، نحصل على:

$$C + \Delta C = B_0 + B_1 (Y + 1)$$

$$C + \Delta C = B_0 + B_1 Y + B_1$$

نعوض عن قيمة C بما يعادلها:

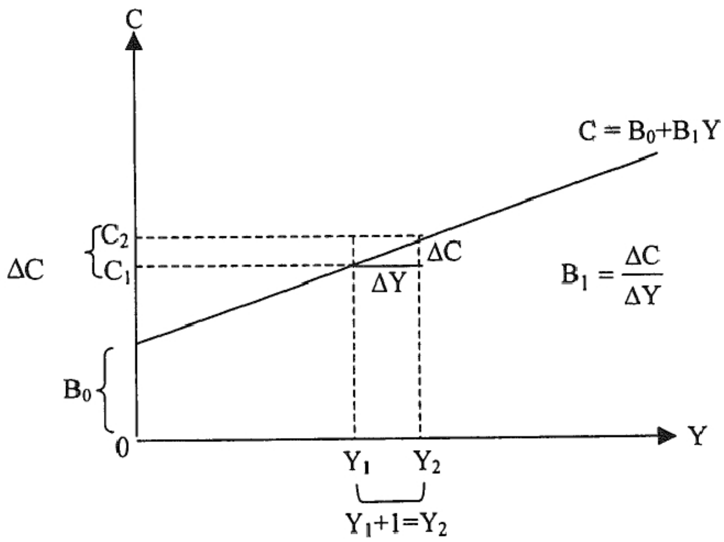
$$B_0 + B_1 Y + \Delta C = B_0 + B_1 Y + B_1$$

وبعد الاختصار، نحصل على: $\Delta C = B_1$

وفي ضوء ذلك يمكن القول أن الميل الحدي للاستهلاك يمثل مقدار الزيادة الحاصلة في C بنسبة زيادة Y بمقدار وحدة واحدة، أي أن: $B_1 = \frac{\Delta C}{\Delta Y}$. وبما أن $\Delta Y = 1$، إذًا: $\therefore\ B_1 = \Delta C$

ب- بيانيــًا:

شكل (2.1)



وتأسيسًا على ذلك هناك عدة خصائص مرغوب فيها لأي نموذج اقتصادي منها:

1- مطابقته للنظرية الاقتصادية، بحيث يصف الظاهرة الاقتصادية بشكل صحيح (نظير مدكور، 2007 صفحة 6).

2- قدرته على توضيح المشاهدات الواقعية، بحيث يكون متناسقًا مع السلوك الفعلي للمتغيرات الاقتصادية التي تحدد العلاقة بين هذه المتغيرات.

3- دقته في تقدير المعلمات، أن هذه التقديرات يجب أن تكون أفضل تقريب للمعلمات الحقيقية. وتأتي هذه الدقة من اتصاف هذه التقديرات بصفات مرغوبة يحددها الاقتصاد القياسي مثل خاصية عدم التحيز Unbaise، وكفاءتها.

4- قدرة النموذج الاقتصادي على التنبؤ، بحيث يعطي تنبؤات مرضية للقيم المستقبلية للمتغيرات المعتمدة.

5- خاصية البساطة، إذ أن النموذج الاقتصادي يجب أن يبرز العلاقات الاقتصادية بأقصى حد ممكن من البساطة، فكلما قل عدد المعادلات وكان شكلها الرياضي أبسط اعتبر النموذج الاقتصادي أفضل من غيره شريطة أن لا يكون على حساب الدقة في التقدير.

5- أنواع النماذج:

هناك عدة أنواع من النماذج التي يمكن تصنيفها كالآتي:

1.5 النماذج الاقتصادية الكلية والجزئية:

أ-**النماذج الاقتصادية الكلية:** وهي النماذج التي تتعامل مع المتغيرات الاقتصادية التي تخص الاقتصاد الكلي أي تتصل بالسلوك العام والبنية العامة للاقتصاد كالدخل، الاستثمار العام...الخ (بخيت، وآخرون، 2010 صفحة 11).

ب- **النماذج الاقتصادية الجزئية:** وهي النماذج التي تتعامل مع المتغيرات الاقتصادية التي تخص الوحدات الاقتصادية الجزئية كعلاقة العرض والطلب على سلعة معينة.

2.5 النماذج الاقتصادية الساكنة والمتحركة:

أ- **النماذج الاقتصادية الساكنة:** وهي النماذج التي لا يكون الزمن أحد متغيراتها أو مؤثرًا في تغيير قيم أحد المتغيرات الداخلة فيها، أي بدون فترة ارتداد زمني، وهذا يعني أن لكل متغير قيمة معينة في السنة التي يقع فيها، فمثلاً تكون دالة الطلب الساكنة كالآتي: $\mathrm{D}_t = \mathrm{f}(\mathrm{P}_t)$

ب-**النماذج الاقتصادية:** وهي النماذج التي يكون الزمن أحد متغيراتها أو مؤثرًا في أحد متغيراتها، إن هذه النماذج توضح كيفية تأثير الزمن في المتغيرات الاقتصادية، وتعد هذه النماذج أكثر واقعية، فمثلاً تكون دالة العرض المتحركة كالآتي: $\mathrm{S}_t = \mathrm{f}(\mathrm{P}_{t-1})$

أي أن العرض في السنة الحالية (t) تعتمد على سعر السلعة في السنة السابقة $(1-t)$ ويسمى المتغير الحركي بمتغير مرتد زمنيًا مثل (P_{t-1}).

6. مكونات النموذج:

1.6 معادلات النموذج:

يتكون النموذج الاقتصادي من مجموعة من المعادلات تسمى بالمعادلات الهيكلية لأنها توضح الهيكل الأساس للنموذج المراد بناؤه، وتختلف عدد المعادلات من نموذج لآخر تبعًا لنوع النموذج والهدف من بنائه. وتنقسم المعادلات الهيكلية إلى:

- **المعادلات السلوكية:** هي المعادلات التي تعبر عن العلاقات الدالية بين المتغيرات الاقتصادية، ويمكن التعبير عنها بدالة ذات متغير مستقل واحد أو عدة متغيرات مستقلة كما يلي: $Y = \beta_0 + \beta_1 X + \varepsilon$

حيث: Y هو المتغير التابع (السلوك المراد تفسيره)، X هو المتغير المستقل (المؤثر على السلوك)، β_0 الثابت، ε الخطأ العشوائي

- **المعادلات التعريفية أو المتطابقات:** هي المعادلات التي تعبر عن علاقة اقتصادية ناتجة عن تعاريف متفق عليها أو هي العلاقة التي تحدد قيمة المتغير التابع بتحديد تعريف له في صورة علاقة مساواة.

2.6 متغيرات النموذج:

تتكون معادلات النموذج من عدد من المتغيرات يمكن تصنيفها إلى عدة أنواع وكما يأتي:

- **المتغيرات الداخلية:** وهي المتغيرات التي تؤثر في النموذج وتتأثر به، وتتحدَّد قيمتها من داخل النموذج عن طريق المعاملات وقيم المتغيرات الخارجية، وتسمى هذه المتغيرات (الداخلية) أيضًا بالمتغيرات التابعة.
- **المتغيرات الخارجية:** وهي المتغيرات التي تؤثر في النموذج ولا تتأثر به، وتتحدَّد قيمتها بعوامل خارجة عن النموذج وفي بعض الأحيان تتحدَّد قيمها عن طريق نموذجٍ آخر مختلف عن النموذج الأصلي وتسمى هذه المتغيرات (الخارجية) بالمتغيرات المستقلة.
- **المتغيرات المرتدة زمنيًا:** وهي المتغيرات التي تنتمي إلى فترة زمنية سابقة أو التي تؤخذ قيمها من الفترة السابقة مثل (Y_{t-1}) التي تمثل دخل السنة الماضية.

7. منهجية الاقتصاد القياسي Methodology of Economics:

يهتم الاقتصاد القياسي بقياس معاملات Coefficients، النموذج المستخدم في التقدير والتنبؤ لقيم المتغيرات الاقتصادية، وهذا يتطلب إتباع منهجية معينة في البحث، لأن العلاقة بين المتغيرات الاقتصادية سببية Causal، أي بمعنى أن التغير في بعض المتغيرات يُحدث أثرًا في المتغيرات الأخرى، ويمكن تحديد هذه المنهجية بالخطوات الآتية:

1.7 مرحلة التوصيف Specification Stage:

تعد مرحلة توصيف (صياغة النموذج من أهم مراحل بناء النموذج وأصعبها وذلك من خلال ما تتطلبه من تحديد للمتغيرات التي يجب أن يشتمل عليها النموذج أو التي يجب استبعادها منه (شاكر، وآخرون، 2006 صفحة 17). وفي هذه المرحلة يتم الاعتماد على النظرية الاقتصادية الاقتصاد الرياضي لتحويل العلاقة المذكورة إلى معادلات رياضية باستخدام الرموز في تحديد نوع واتجاه العلاقة بين المتغيرات الاقتصادية، كما يتم الاعتماد على الرياضية مثل العلاقة بين الكمية المطلوبة من سلعة ما (D_x) والسعر (P_x) والدخل (Y) حيث تصاغ العلاقة أعلاه كالآتي:

$$D_x = B_o + B_1 P_x + B_2 Y \dots\dots(3.1)$$

فمن نظرية الطلب يتوقع الحصول على إشارة سالبة للمعامل B_1 وذلك لوجود علاقة عكسية بين الكمية المطلوبة من سلعة ما وسعرها حسب النظرية الاقتصادية وإشارة موجبة للمعامل B_2 لوجود علاقة طردية بين الكمية المطلوبة ودخل المستهلك، كما يتم جمع البيانات الخاصة بمتغيرات النموذج.

2.7 مرحلة التقدير Estimation Stage:

في هذه المرحلة يتم جمع البيانات المتعلقة بالظاهرة الاقتصادية (المشكلة) قيد الدراسة، ومن ثم يتم تقدير معالم العلاقة التي تم وصفها وصياغتها رياضيًا في المرحلة الأولى، أي تقدير قيم رقمية للمعالم B_o، B_1، B_2 في دالة الطلب أعلاه كما يجب في هذه المرحلة تقييم المعالم المقدرة من النواحي الاقتصادية والإحصائية والقياسية.

فمن الناحية الاقتصادية تجري عملية مقارنة بين قيم وإشارات معالم النموذج لتي تم تقديرها مع القيم والإشارات المتوقعة لهذه المعالم في ضوء النظرية الاقتصادية.

ومن الناحية الإحصائية يتم حساب الانحرافات الكلية والجزئية في المتغيرات التي يتكون منها النموذج واختبار معنوية المعالم من خلال اختبار (t) ومعامل التحديد (R_2).

أما من الناحية القياسية فيتم اختبار مدى انسجام وتحقق الفروض الخاصة بالمتغير العشوائي على النموذج القياسي المقترح، حيث أن وجود الاختلاف يعني وجود مشاكل منها مشكلة الارتباط الذاتي، التعدد الخطي، وعدم ثبات تجانس التباين والتي سيتم التعرف على كلٍ منها بشكل مفصل وبفصل خاص لكل منها في الفصول اللاحقة.

3.7 مرحلة الاختبار Testing Stage:

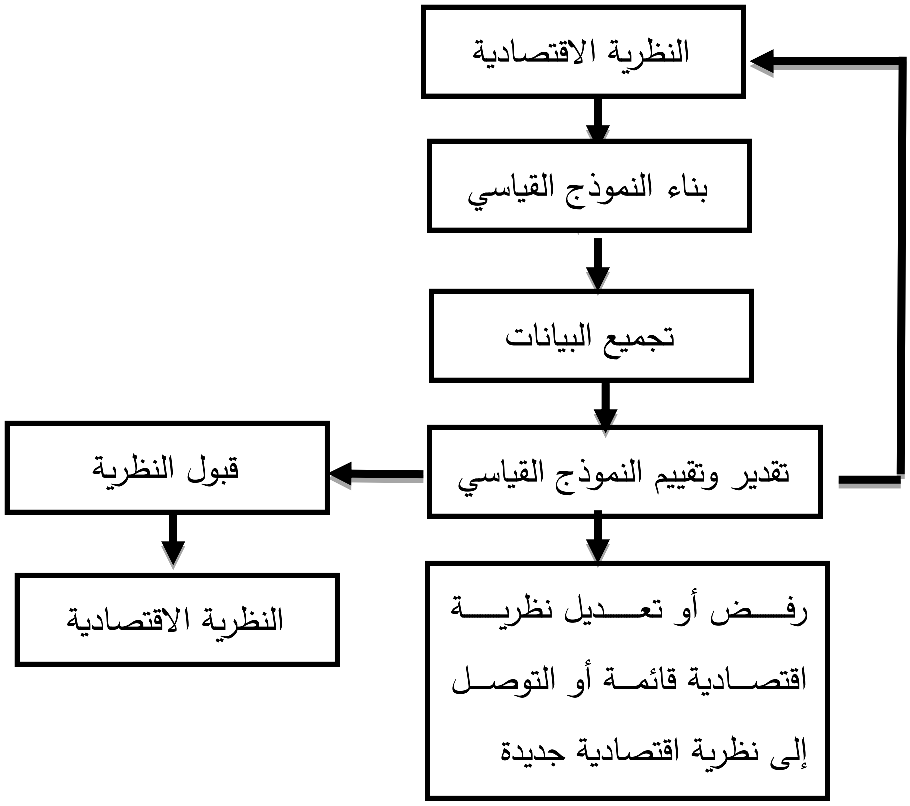
في هذه المرحلة يتم اختبار قوة ومعنوية النموذج المقدر باعتماد طرق إحصائية معينة للتأكد من صلاحية النموذج وقدرته على التنبؤ . وقد يواجه الباحث هنا عدة مشاكل، منها مشكلة تغاير حد الخطأ أو الارتباط الذاتي أو الازدواج الخطي وغيرها من المشاكل، وعلى الباحث أن يعالج هذه المشاكل قبل البدء بعملية التقييم.

4- مرحلة التنبؤ، Prediction Stage:

لا يوجد من يعترض على ضرورة التنبؤ بالمستقبل والتعرف عليه مسبقًا قبل قدومه وعلى مختلف المستويات الكلية والجزئية –وفي مختلف المجالات الاقتصادية والاجتماعية ولمختلف المدد القصيرة والمتوسطة والطويلة عليه يتم في هذه المرحلة أعداد تقديرات مستقبلية للمتغيرات المدروسة كحجم الطلب على السلعة (D_x) في مثالنا السابق.

ولكن قبل استخدام النموذج المقدر في التنبؤ يجب التأكد من جودة الأداء العام للنموذج المقدر، وبعدئذ يتم تطبيق النتائج التي تم التوصل إليها على الواقع واستخدامها في عملية التنبؤ. ويمكن توضيح منهجية البحث في الاقتصاد القياسي كما مبين في الشكل (2.1).

الشكل (3.1) : منهجية البحث في الاقتصاد القياسي.



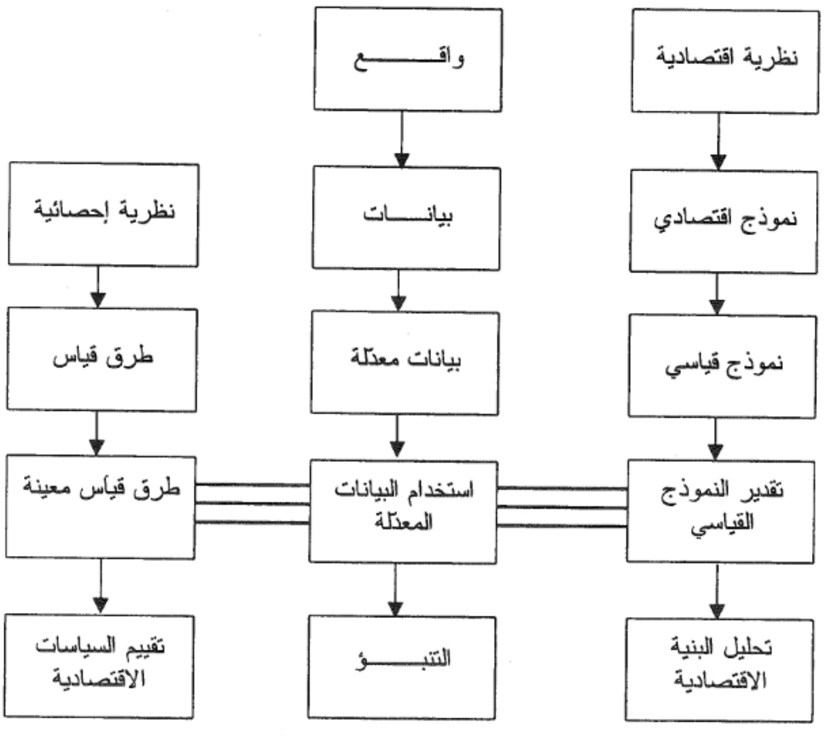
8- أسلوب الاقتصاد القياسي Econometrics Approach:

يعتمد أسلوب الاقتصاد القياسي على النظرية الاقتصادية التي تزوده بالنماذج الاقتصادية التي تستخدم في تكوين النماذج القياسية، ولغرض تقدير أو تقييم النماذج القياسية ووضعها في شكل كمي، يعتمد الاقتصاد القياسي على الواقع الذي تستقي منه المعلومات والبيانات اللازمة لعملية التقدير ومن ثم يقوم بعملية التصحيح أو تعديل البيانات المستقاة من الواقع وتهيئتها للاستخدام.

وبعدها يعود الاقتصاد القياسي إلى نظرية الإحصاء بغرض تطوير واستنساخ طرق القياس وبذلك يكون تحت التصرف نموذج قياسي مع بيانات صالحة للاستخدام مع طرق القياس ومن ثم يتم استخدام هذه الإمكانيات في تقدير معلمات أو معاملات النموذج القياسي الذي يجسد العلاقات الاقتصادية.

ومن ثم يمكن استخدام هذا النموذج الذي تم تقديره، إما في تحليل البنية الاقتصادية أو في التنبؤ أو في تقييم السياسات الاقتصادية، ويمكن عرض ذلك بشكل توضيحي من خلال الشكل (4.1)

الشكل (4.1): أسلوب الاقتصاد القياسي



المحور الثاني: تحليل الانحدار الخطي البسيط

(تحديد قيم معلمات النموذج؛ اختبار الموثوقية، التنبؤ)

1. مقدمــة:

إن دراسة العلاقة بين المتغيرات الاقتصادية يتطلب تحديد المتغيرات المؤثرة في تلك العلاقة ومن أبسط وأسهل أنواع العلاقات في التقدير والتحليل الإحصائي والاقتصادي، العلاقة بين متغيرين أحدهما المتغير التابع Dependent Variable والثاني المتغير المستقل Independent Variable ، وإذا رمزنا للمتغيرين بـY و X وعلى التوالي، فإن العلاقة الدالية التي تجمعهما تكون كالآتي:

$$Y = F(x) \ldots\ldots (1.2)$$

حيث يشير الرمز F إلى كون المتغير التابع Y يعتمد على المتغير المستقل X. ولتحديد شكل العلاقة هذه – ما إذا كان خطياً أم غير خطي – يمكن الاستعانة بالنظرية الاقتصادية، كما يستعان بالاقتصاد الرياضي والإحصاء لصياغة العلاقة واختبار المتغيرات، كما لابد من تحديد شكل العلاقة هذه إذ تحكم العلاقة بين المتغيرات بعدد من الأشكال (الصيغ) أبسطها وأكثرها شيوعًا الصيغة الخطية، وتسمى العلاقة الخطبة بين متغيرين بالانحدار الخطي البسيط ، Simple Linear Regression، فالعلاقة الخطية بين X ولتكن دخل الأسرة و Y ولتكن الإنفاق على سلعة معينة يمكن أن تكتب بالصيغة الرياضية الآتية:

$$Y_i = B_0 + B_1 X_i \ldots\ldots\ldots (2.2)$$

حيث B_1 وB_0 عبارة عن معلمات مجهولة القيم وثوابت يُشرحان من وجهة النظر الرياضية كالآتي:

B_0 : تمثل تقاطع خط الانحدار مع المحور العمودي وهي عبارة عن القيمة التي تتخذها Y عندما تكون قيمة X مساوية للصفر.

B_1: تمثل الميل.

ومن وجهة النظر الاقتصادية تمثل (B_0) حالة الكفاف و(B_1) الميل الحدي للاستهلاك، وقيمة الميل عبارة عن مقدار الزيادة المتحققة في قيمة المتغير التابع Y نتيجة زيادة المتغير المستقل بمقدار وحدة واحدة.

غير أن العلاقة أعلاه (2.2) لا يمكن أن تشرح العلاقة بين المتغيرين بشكل دقيق، فهناك أسباب مهمة تجعل هذه المعادلة غير معبرة عن العلاقة بين X و Y تعبيراً كاملاً فقد يكون هناك انحراف بين العلاقة الحقيقية والمعادلة الإحصائية التي تمثلها نتيجة أخطاء في القياسات أو في اختيار المتغير المستقل، مما يتطلب إضافة متغير جديد يسمى بالحد العشوائي Random Variable ويرمز له عادة بالرمز (U) ودوره امتصاص العوامل غير القابلة للقياس، وكذلك أخطاء القياس، عليه فإن العلاقة من الصيغة (2.2) يجب أن تعدل لكي تضم حد الخطأ العشوائي حيث يصبح:

$$\mathrm{Y}_i = \mathrm{B}_0 + \mathrm{B}_i \mathrm{X}_i + U_i \ldots\ldots (3.2)$$

2. الفرضيات الخاصة بالمتغير العشوائي:

1 –إن المتغير العشوائي (U_i) هو متغير تعتمد قيمته في أية فترة زمنية على عامل الصدفة (بخيت علي، 2009 صفحة 30)، فقد تكون أكبر أو أصغر أو مساوية إلى الصفر، إلا أنها في المتوسط تساوي صفر، أي $E(U_i) = 0$، ويمكن توضيح ذلك على النحو الآتي:

$$Y_i = B_0 + B_1X_i + U_i \quad \ldots\ldots\ldots(4.2)$$
$$U_i = Y_i - B_0 - B_1X_i \ldots\ldots\ldots(5.2)$$

وبإدخال Σ على طرفي المعادلة 5.2:

$$\sum U_i = \sum (Y_i - B_0 - B_1X_i)$$
$$\sum U_i = \sum Y_i - n\,B_0 - B_1 \sum X_i \ldots\ldots(6.2)$$
$$\therefore B_0 = \bar{Y} - B_1\bar{X}$$

نعوض عن B_0 بما يساويها في المعادلة (6.2) :

$$\sum U_i = \sum Y_i - n(\bar{Y} - B_1\bar{X}) - B_1 \sum X_i \ldots\ldots(7.2)$$
$$\therefore \ \bar{Y} = \frac{\sum Y_i}{n} \ , \ \ \bar{X} = \frac{\sum X_i}{n}$$

وبحاصل ضرب الطرفين في الوسطين نحصل:

$$\sum Y_i = n\bar{Y} \quad , \qquad \sum X_i = n\bar{X}$$

وبالتعويض عن ذلك في المعادلة (7.2) تكون:

$$\sum U_i = \sum Y_i - \sum Y_i + B_1 \sum X_i - B_1 \sum X_i$$
$$\sum U_i = 0$$
$$E(U_i) = 0$$

2– إن المتغير العشوائي (U_i) يتوزع توزيعاً طبيعياً ، Normally distributed حول القيمة المتوقعة أو حول الوسط الحسابي المساوي للصفر عند كل قيمة من قيم المتغير المستقل X أي بشكل جرس.

3–إن تباين Variance، المتغير العشوائي (حد الخطأ)، حول الوسط الحسابي مقدار ثابت عند كل قيمة من قيم X أي:

$$var(U_i) = E[U_i - E(U_i)^2]$$
$$\therefore E(U_i) = 0$$
$$\therefore var(U_i) = E(U_i)^2 = \sigma^2$$

وإذا كـان تبـاين الخطـأ غيـر ثابـت عندئـذ تظهـر مشـكلة تسـمى مشـكلة عـدم تجـانس التبـاين Heteroscedasticity، والتي سنتناولها بشيء من التفصيل لاحقاً.

الفرضيات الثلاث السابقة يمكن جمعها بشكل مختصر وتمثيلها كالآتي:

$$U_i \sim N(0, \sigma^2)$$

أي بمعنى أن الخطأ العشوائي، U_i، يتوزع ~ توزيعاً طبيعياً، N، بوسط حسابي مساوي للصفر، 0، وتباين ثابت قيمته σ^2.

4- أن قيم U_i غير مرتبطة بأي من المتغيرات المستقلة، أي انعدام التباين المشترك Covariance بين U_i و X_i أي:

$$Cov(U_iX_i) = \mathrm{E}(U_iX_i)$$
$$Cov(U_iX_i) = X_iE(U_i)$$
$$\therefore\ E(U_i) = 0$$
$$\therefore E(U_iX_i) = 0$$

ويمكن توضيح ذلك على النحو الآتي:

$$Y_i = B_0 + B_1X_i + U_i \ldots\ldots\ldots.(8.2)$$
$$U_i = Y_i - B_0 - B_1X_i.\ldots\ldots.(9.2)$$

وبضرب طرفي المعادلة بـ $\sum X_i$:

$$\sum X_i\, U_i = \sum X_i\, Y_i - B_0 \sum X_i - B_1 \sum X_i^2$$
$$\therefore\ B_0 = \bar{Y} - B_1\bar{X}$$

وعند تعويض ذلك :

$$\sum X_iU_i = \sum X_iY_i - \sum X_i\,(\bar{Y} - B_1\bar{X}) - B_1 \sum X_i^2 \ldots\ldots\ldots(10.2)$$
$$\therefore \bar{Y} = \frac{\sum Y_i}{n} \quad , \quad \bar{X} = \frac{\sum X_i}{n}$$

وبتعويض ذلك يكون:

$$\sum X_iU_i = \sum X_iY_i - \sum X_i\left(\frac{\sum Y_i}{n} - B_1\frac{\sum X_i}{n}\right) - B_1\sum X_i^2$$
$$\sum X_iU_i = \sum X_iY_i \sum X_i\left(\frac{\sum Y_i - B_1\sum X_i}{n}\right) - B_1\sum X_i^2$$
$$\sum X_iU_i = \sum X_iY_i \sum X_iY_i + B_1\sum X_i^2 - B_1\sum X_i^2$$

وبعد الحذف والتبسيط يكون:

$$\sum X_i U_i = 0$$

5- القيم المختلفة للمتغير العشوائي (U_i) تكون مستقلة عن بعضها البعض، بعبارة أخرى التباين المشترك لـ U_i مع U_j مساوٍ للصفر، وعليه فإن قيمة العنصر العشوائي في أي فترة لا تعتمد على قيمته في فترة أخرى أي:

$$cov\,(U_i U_j) = E(U_i U_j) = 0 \qquad (i, j \;=\; 1,2,3, \ldots, n, i \neq j)$$

وإذا حدث وجود ارتباط بينها تظهر مشكلة تسمى مشكلة الارتباط الذاتي Autocorrelation، وسيتم شرحها لاحقاً (بخيت علي، 2009 صفحة 39).

3. طريقة المربعات الصغرى The Ordinary Least Squares (OLS):

بالرجوع إلى العلاقة الخطية بين دخل الأسرة X وإنفاقها على سلعة معينة Y:

$$Y_i = B_0 + B_1 X_i + U_i \ldots\ldots\ldots\ldots (11.2)$$

يتبين لنا بأن تأثير الدخل في الإنفاق على السلعة موضوع البحث يتحدد من خلال العلاقة المنتظمة ($B_0 + B_1 X_1$)، أما تأثير العوامل الأخرى فإنه متجسد في (U_i).

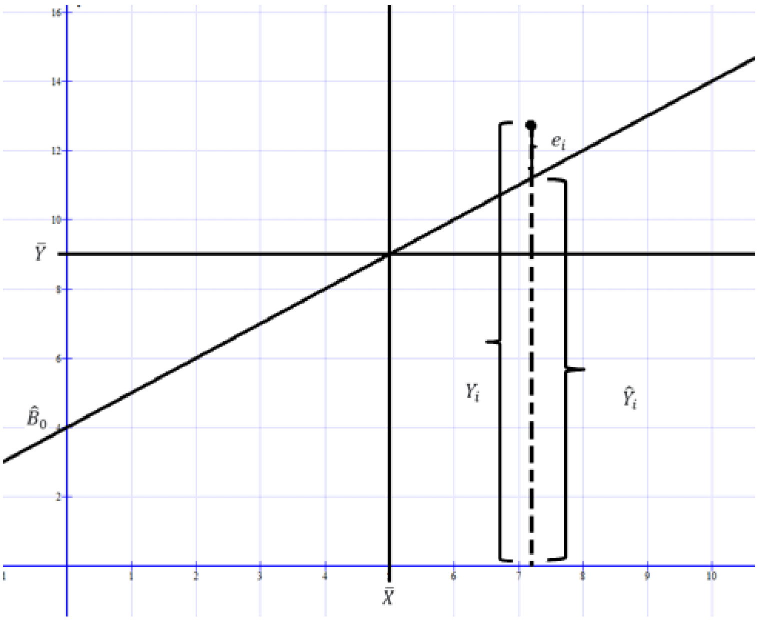
وعليه فإنه لمعرفة العلاقة الحقيقية بين دخل الأسرة وإنفاقها على السلعة في القطر يتطلب احتساب B_0 و B_1 ، إلا أن احتساب المعالم المذكورة لا يمكن أن يتم إلا في حالة الحصول على دخل وإنفاق جميع الأسر في ذلك القطر وهذا أمر غير ممكن بسبب صعوبة العملية الإحصائية اللازمة ولتسهيل العمل تسحب عينة من أسر القطر، ومن ثم تقدر قيم المعالم ويتم التقدير بواسطة المعادلة:

$$Y_i = \hat{B}_0 + \hat{B}_1 X_i + e_i \ldots\ldots\ldots (12.2)$$

ولتقدير تأثير الدخل بصورة مستقلة في الإنفاق فإنه يتم بواسطة المعادلة التالية:

$$\hat{Y}_i = \hat{B}_0 + \hat{B}_1 X_i \ldots\ldots\ldots\ldots (13.2)$$

تسمى المعادلة (13.2) بمعادلة خط الانحدار، وتشير العلامة (^) إلى كون القيم تقديرية وليست حقيقية وكل نقطة من نقاطه (Y_i) تمثل القيمة التقديرية لمتوسط إنفاق جميع العوائل ذات الدخل البالغ X . ويتبين من المعادلتين (12.2) و (13.2) بأن قيم المشاهدات الفعلية Y_i تنحرف عن القيم التقديرية ($\hat{Y}_i$) بمقدار e_i وكما مبين في الشكل الآتي: الشكل (1.2)



من الشكل يتبين أن : $e_i = Y_i - \hat{Y}_i$

حيث يمكن للبواقي e_i أن تكون سالبة أو موجبة حسب موضع نقطة المشاهدة من الخط المقدر . ولإيجاد أفضل خط مستقيم لعينة مشاهدات X ، Y من بين خطوط لا نهائية العدد تصف المعادلة الخطية تستخدم طريقة المربعات الصغرى (OLS)، ويتضمن ذلك في محاولة جعل مجموع مربع انحرافات القيم الحقيقية Y_i عن القيم التقديرية $\hat{Y}_i$ أقل ما يمكن، أي جعل مجموع مربعات الأخطاء العشوائية عند نهايتها الصغرى وبما أن طريقة OLS تشترط تصغير القيمة $(\sum e_i^2)$ إلى الحد الأدنى فإنها عبارة عن مشكلة النهايات الصغرى أي:

$$\min \rightarrow \sum_{i=1}^{n} e_i^2$$

حيث أن: $$e_i = Y_i - \hat{Y}_i$$

إذًا: $$\sum e_i^2 = \sum \left(Y_i - \hat{Y}_i\right)^2$$

بما أن معادلة الخط المستقيم الحقيقية غير المعروفة هي: $Y_i = B_0 + B_1 X_i$

فإن معادلة الخط المستقيم التقديرية تكون: $\hat{Y}_i = \hat{B}_0 + \hat{B}_1 X_i$

بالتعويض عن $\hat{Y}_i$ بما يساويها نحصل:

$$\sum e_i^2 = \sum \left(\hat{Y}_i - \hat{B}_0 - \hat{B}_1 X_i\right)^2$$

وكشرط رياضي لتصغير $\sum e_i^2$ تؤخذ المشتقات الجزئية لكل من $\hat{B}_0$ و $\hat{B}_1$ ومساواة كل منها بالصفر.

وإن الشرط الجوهري للتصغير هو أخذ التفاضل الجزئي لمجموع المربعات بالنسبة لمعاملات Coefficient ، النموذج $\hat{B}_0$و$\hat{B}_1$ ومساواة المشتقة الأولى بالصفر.

أي بتطبيق الشرط الضروري:

$$\frac{\partial \sum e_i^2}{\partial \hat{B}_0} = 2\sum\left(Y_i - \hat{B}_0 - \hat{B}_1 X_i\right)(-1) = 0$$

$$-2\sum\left(Y_i - \hat{B}_0 - \hat{B}_1 X_i\right) = 0$$

بالقسمة على (-2) وفك القوس وترتيب المعادلة نحصل على:

$$\sum\left(Y_i - \hat{B}_0 - \hat{B}_1 X_i\right) = 0$$

$$\sum Y_i - n\hat{B}_0 - \hat{B}_1 \sum X_i = 0$$

$$\sum Y_i = n\hat{B}_0 + \hat{B}_1 \sum X_i \ldots\ldots\ldots.(14.2)$$

$$\frac{\partial \sum e_i^2}{\partial \hat{B}_1} = 2\sum\left(Y_i - \hat{B}_0 - \hat{B}_1 X_i\right)(-X_i) = 0$$

$$-2\sum X_i\left(Y_i - \hat{B}_0 - \hat{B}_1 X_i\right) = 0$$

بالقسمة على (-2) نحصل على :

$$\sum X_i\left(Y_i - \hat{B}_0 - \hat{B}_1 X_i\right) = 0$$

$$\sum X_i Y_i - \hat{B}_0 \sum X_i - \hat{B}_1 \sum X_1^2 = 0$$

$$\sum X_i Y_i = \hat{B}_0 \sum X_i - \hat{B}_1 \sum X_1^2 \ldots\ldots\ldots(15.2)$$

تسمى المعادلتين (14.2) و (15.2) بالمعادلتين الطبيعيتين (الآنيتين) حيث (n) عدد المشاهدات X_i وY_i هي معلومة دائمًا باعتبارهما قيم المشاهدات الحقيقية وبمجرد تعويضهما في المعادلتين (14.2) و (15.2) وبحلهما آنيًا نحصل على قيم $\hat{B}_0$ و$\hat{B}_1$ اللتان تمثلان مقدرتان للمعلمتين الحقيقيتين B_0 و B_1 .

1.3 طرق تقدير معاملات النموذج:

ولتقدير معاملات النموذج B_0 و B_1 نستعين بعدة طرق منها:

1 – طريقة الحذف والتعويض.

2– طريقة المحددات.

3 – طريقة التقدير حول نقطة المتوسط.

4– طريقة المصفوفات.

وسنبين هذه الطرق من خلال المثال الآتي، من دون تكرارها هنا وهناك.

مثال 1.2: الجدول الآتي يمثل عدد سنوات الخدمة (X_I) ومعدل الأجر السنوي (Y_I) بآلاف الدنانير لعينة تمثل (8) موظفين في أحد الدوائر.

X_i	Y_i	X_iY_i	X_i^2
4	25.6	102.4	16
8	32.7	261.6	64
12	45.4	544.8	144
16	53.9	862.4	256
20	59.0	1180	400
24	62.6	1502.4	576
28	65.0	1820	784
32	65.5	2105.6	1024
$\sum X_i = 144$ $\bar{X} = 18$	$\sum Y_i = 410$ $\bar{Y} = 51.25$	$\sum X_iY_i = 8379.2$	$4\sum X_i^2 = 326$

المطلوب : تقدير خط الانحدار بواسطة المعادلتين الطبيعيتين أعلاه.

1.1.3 طريقة الحذف والتعويض Substitution Method:

$$\sum Y_i = n\hat{B}_0 + \hat{B}_1 \sum X_i \ldots\ldots\ldots.(16.2)$$
$$\sum X_iY_i = \hat{B}_0 \sum X_i + \hat{B}_1 \sum X_i^2 \ldots\ldots(17.2)$$

وبتعويض القيم من الجدول في المعادلتين 16.2 ، 17.2 نحصل على:

$$410 = 8\hat{B}_0 + 144\hat{B}_1 \ldots\ldots\ldots\ldots\ldots.(18.2)$$
$$8379.2 = 144\hat{B}_0 + 3264\hat{B}_1 \ldots\ldots.(19.2)$$

وبضرب المعادلة (18.2) في 18 نحصل:

$$7380 = 144\hat{B}_0 + 2592\hat{B}_1 \ldots\ldots\ldots\ldots\ldots.(20.2)$$
$$8379.2 = 144\hat{B}_0 + 3264\hat{B}_1 \ldots\ldots\ldots\ldots\ldots.(21.2)$$

وبطرح المعادلة (20.2) من (21.2) نحصل:

$$999.2 = 672\hat{B}_1$$
$$\hat{B}_1 = \frac{999.2}{672} = 1.486904762$$

للحصول على قيمة $\hat{B}_0$ نعوض عن قيمة $\hat{B}_1$ في أحد المعادلتين الرئيستين ولتكن معادلة (18.2):

$$410 = 8\hat{B}_0 + 144\hat{B}_1$$
$$410 = 8\hat{B}_0 + 144(1.486904762)$$
$$410 = 8\hat{B}_0 + 214.1142857$$

$$410 - 214.1142857 = 8\hat{B}_0$$

$$195.8857143 = 8\hat{B}_0$$

$$\hat{B}_0 = \frac{195.8857143}{8} = 24.48571429$$

وعليه فإن المعادلة المقدرة Estimated، للعلاقة بين عدد سنوات الخدمة X_i ومعدل الأجر السنوي Y_i للعينة المعنية تكون:

$$\hat{Y}_i - \hat{B}_0 + \hat{B}_1 X_i$$

$$\hat{Y}_i = 24.48571429 + 1.486904762 X_i$$

تشير المعادلة التقديرية إلى وجود علاقة طردية بين المتغير التابع y_i الذي يمثل معدل الأجر السنوي للموظف والمتغير المستقل X_i الذي يمثل عدد سنوات الخدمة، فبزيادة خدمته الوظيفية بمقدار سنة واحدة يزداد معدل أجره السنوي بمقدار 1486 دينار.

2.1.3. طريقة المحددات Determinates Method:

ويمكن الحصول على قيم $\hat{B}_0$ و $\hat{B}_i$ باعتماد المحددات (قاعدة كرايمر) وذلك بإعادة كتابة المعادلتين الطبيعيتين (16.2) و (17.2) في صيغة مصفوفة وعلى النحو الآتي:

$$\sum Y_i = n\hat{B}_0 + \hat{B}_1 \sum X_i$$

$$\sum X_i Y_i = \hat{B}_0 \sum X_i + \hat{B}_1 \sum X_i^2$$

$$\begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \begin{bmatrix} \hat{B}_0 \\ \hat{B}_1 \end{bmatrix}$$

ولتقدير $\hat{B}_0$ و $\hat{B}_1$ ينبغي تكوين المحددات الآتية:

$$|D| = \begin{vmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{vmatrix} = (n)\left(\sum X_i^2\right) - \left(\sum X_i\right)\left(\sum X_i\right)$$

$$|A_0| = \begin{vmatrix} \sum Y_i & \sum X_i \\ \sum X_i Y_i & \sum X_i^2 \end{vmatrix} = \left(\sum Y_i\right)\left(\sum X_i^2\right) - \left(\sum X_i Y_i\right)\left(\sum X_i\right)$$

$$|A_1| = \begin{vmatrix} \text{n} & \sum Y_i \\ \sum X_i & \sum X_i Y_i \end{vmatrix} = (\text{n})\left(\sum X_i Y_i\right) - \left(\sum X_i\right)\left(\sum Y_i\right)$$

$$\hat{B}_1 = \frac{|A_1|}{|D|} = \frac{n(\sum X_i Y_i) - (\sum Y_i)(\sum Y_i)}{n(\sum X_i^2) - (\sum X_i)^2}$$

$$\hat{B}_0 = \frac{|A_0|}{|D|} = \frac{(\sum Y_i)(\sum X_i^2) - (\sum X_i Y_i)(\sum X_i)}{n(\sum X_i^2) - (\sum X_i)^2}$$

$$or: \quad \hat{B}_0 = \bar{Y} - \hat{B}_1 \bar{X}$$

بالرجوع إلى بيانات المثال (1.2) وباعتماد المحددات نحصل على قيم $\hat{B}_0$ و$\hat{B}_1$ وكما يلي:

$$\begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix} \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \begin{bmatrix} \hat{B}_0 \\ \\ \hat{B}_1 \end{bmatrix}$$

$$|D| = \begin{vmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{vmatrix} = \begin{vmatrix} 8 & 144 \\ 144 & 3264 \end{vmatrix}$$

$$|D| = (8)(3264) - (144)(144)$$

$$|D| = 26112 - 20736$$

$$|D| = 5376$$

$$|A_0| = \begin{vmatrix} \sum Y_i & \sum X_i \\ \sum X_i Y_i & \sum X_i^2 \end{vmatrix} = \begin{vmatrix} 410 & 144 \\ 8379.2 & 3254 \end{vmatrix}$$

$$|A_0| = (410)(3264) - (144)(9379.2)$$

$$|A_0| = 1338240 - 1206604.8$$

$$|A_0| = 131635.2$$

$$|A_1| = \begin{vmatrix} \text{n} & \sum Y_i \\ \sum X_i & \sum X_i Y_i \end{vmatrix} = \begin{vmatrix} 8 & 410 \\ 144 & 8379.2 \end{vmatrix}$$

$$|A_1| = (8)(8379.2) - (410)(144)$$

$$|A_1| = 67033.6 - 59040$$

$$|A_1| = 7993.6$$

$$\hat{B}_0 = \frac{|A_0|}{|D|} = \frac{131635.2}{5376} = 24.48571429$$

$$\hat{B}_1 = \frac{|A_1|}{|D|} = \frac{7993.6}{5376} = 1.486901762$$

وبذلك تكون المعادلة التقديرية:

$$\therefore \hat{Y}_i = 24.4857 + 1.486X_i$$

3.1.3. طريقة التقدير حول نقطة المتوسط:

كما يمكن تقدير $\hat{B}_0$ و $\hat{B}_1$ بواسطة انحرافات المتغيرين Y_i و X_i عن وسطهما الحسابي $\bar{Y}$ و $\bar{X}$ باستخدام فكرة البواقي e_i فبقسمة المعادلة الطبيعية رقم (1) على n نحصل على:

$$\sum Y_i = n\hat{B}_0 + \hat{B}_1 \sum X_i \ldots \ldots .(22.2)$$
$$\frac{\sum Y_i}{n} = \hat{B}_0 \frac{n}{n} + \hat{B}_1 \frac{\sum X_i}{n}$$
$$\bar{Y} = \hat{B}_0 + \hat{B}_1 \bar{X} \ldots \ldots \ldots .(23.2)$$
$$\therefore \hat{B}_0 = \bar{Y} - \hat{B}_1 \bar{X} \ldots \ldots (24.2)$$

ولإيجاد $\hat{B}_1$ نعود إلى معادلة الخط المستقيم التقديرية حيث:

$$\hat{Y}_i = \hat{B}_0 + \hat{B}_1 X_i \ldots \ldots \ldots \ldots (25.2)$$

وبطرح المعادلة رقم (23.2) من (25.2) نحصل على:

$$\hat{Y}_i - \bar{Y} = \hat{B}_0 + \hat{B}_1 X_i - \hat{B}_0 - \hat{B}_1 \bar{X} \ldots \ldots \ldots ..(26.2)$$
$$\hat{Y}_i - \bar{Y} = \hat{B}_1 X_i - \hat{B}_1 \bar{X}$$
$$\hat{Y}_i - \bar{Y} = \hat{B}_1 (X_i - \bar{X})$$
$$\hat{y}_i = \hat{B}_1 X_i \ldots \ldots \ldots \ldots .(27.2)$$
$$\because e_i = y_i - \hat{y}_i$$
$$e_i = y_i - \hat{B}_1 x_i \ldots \ldots \ldots (28.2)$$

وبإدخال Σ وتربيع الطرفين:

$$\sum e_i^2 = \sum \left(y_i - \hat{B}_1 x_i\right)^2 \ldots \ldots \ldots \ldots (29.2)$$

بإيجاد المشتقة الجزئية لـ $\hat{B}_1$ ومساواتها بالصفر:

$$\frac{\partial \sum e_i^2}{\sum \hat{B}_1} = 2 \sum \left(y_i - \hat{B}_1 x_i\right)(-x_i) = 0 \ldots \ldots \ldots .(30.2)$$
$$-2 \sum x_i \left(y_i - \hat{B}_1 x_i\right) = 0$$

وبالقسمة على (2−):

$$\sum x_i \left(y_i - \hat{B}_1 x_i\right) = 0$$
$$\sum x_i y_i - \hat{B}_1 \sum x_i^2 = 0$$
$$\sum x_i y_i = \hat{B}_1 \sum x_i^2$$

حيث أن $\sum x_i y_i$ تمثل مجموع حاصل ضرب انحرافات القيم عن الوسط الحسابي للمتغيرين X_i و Y_i وأن $\sum x_i^2$ تمثل مجموع مربع الانحرافات لقيم المتغير X_i عن وسطه الحسابي. وعليه فإن المعادلتين (24.2) و(31.2) هي المعادلات الأساسية التي تستخدم في إيجاد قيم $\widehat{B}_0$ و $\widehat{B}_1$ بموجب طريقة الانحرافات.

وبالرجوع إلى بيانات الجدول (1.2) والتعبير عنها بصيغة انحرافات يمكن الحصول على قيم $\widehat{B}_0$ و $\widehat{B}_1$ كما يأتي:

X_i	Y_i	x_i	Y_i	$x_i y_i$	x_i^2
4	25.6	-14	-25.65	359.1	196
8	32.7	-10	-18.55	185.5	100
12	45.4	-6	-5.85	35.1	36
16	53.9	-2	2.65	-5.3	4
20	59.0	2	7.75	15.5	4
24	62.6	6	11.35	68.1	36
28	65.0	10	13.75	137.5	100
32	68.5	14	14.55	203.7	196
$\sum X_i = 144$ $\bar{X} = 18$	$\sum Y_i = 410$ $\bar{Y} = 51.25$	$\sum x_i = 0$ $x_i = X - \bar{X}$	$\sum y_i = 0$ $y_i = Y - \bar{Y}$	$\sum x_i y_i = 999.2$	$\sum x_i^2 = 672$

$$\hat{B}_1 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{999.2}{672} = 1.486904762$$

$$\hat{B}_1 = \bar{Y} - \hat{B}_1 \bar{X}$$

$$\hat{B}_0 = 51.25 - (1.486904762)(18)$$

$$\hat{B}_0 = 51.25 - 26.76428572$$

$$\hat{B}_0 = 24.48571428$$

$$\therefore \hat{Y}_i = 24.485 + 1.486\, X_i$$

4.1.3 طريقة المصفوفات Matrices Method:

إضافة إلى الطرق السابقة فإنه يمكن تقدير قيم $\hat{B}_0$ و $\hat{B}_1$ باعتماد صيغة المصفوفات حيث يمكن كتابة المعادلات الطبيعية على شكل مصفوفات Matrices ومتجهات Vectors، وكما يلي (بخيت علي، 2009 صفحة 50):

نفترض أن هناك علاقة تحتوي على i من المشاهدات من $1 \leftarrow n$ وأن هناك عدد من المتغيرات من $1 \leftarrow K$.

$$Y_i = B_0 + B_1 X_i + U_i \ldots\ldots\ldots\ldots\ldots. (32.2)$$

$$Y_1 = B_0 + B_1 X_{11} + B_2 X_{12} + \cdots .. + B_k X_{1K} + U_1$$

$$Y_1 = B_0 + B_1X_{11} + B_2X_{12} + \cdots..+B_kX_{1K} + U_1$$
$$Y_1 = B_0 + B_1X_{21} + B_2X_{22} + \cdots..+B_kX_{2K} + U_2$$
$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$
$$Y_n = B_0 + B_1X_{n1} + B_2X_{n2} + \cdots..+B_kX_{nK} + U_n$$

يمكن تمثيل هذه المعادلات بصيغة رمز المصفوفات:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots X_{1k} \\ 1 & X_{21} & X_{22} & \cdots X_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots X_{nk} \end{bmatrix} \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_k \end{bmatrix} + \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix}$$

$$Y = XB + U \ldots\ldots..(33.2)$$

حيث أن:

Y : متجه عمودي أبعاده (nx1) يحتوي على n مشاهدة للمتغير التابع Y.

X: مصفوفة أبعادها $(n \times k + 1)$ تحتوي مشاهدات المتغيرات المستقلة X. وعمودها الأول يحتوي على قيم الواحد. صحيح لأخذ الثابت بنظر الاعتبار.

B: منحه عمودي أبعاده $(k + 1 \times 1)$ تحتوي المعالم المجهولة.

U: متجه عمودي أبعاده (nx1) يحتوي الخطأ العشوائي.

وللحصول على تقديرات المربعات الصغرى العادية لمتجه المعلمات يمكن كتابة المعادلة المقدرة التي يراد الحصول عليها وبصيغة المصفوفات كما يلي:

$$\mathrm{Y} = \hat{\mathrm{Y}} + \mathrm{e}$$
$$\mathrm{Y} = \mathrm{X}\hat{\mathrm{B}} + \mathrm{e}$$
$$\therefore \mathrm{e} = \mathrm{Y} - \mathrm{X}\hat{\mathrm{B}} \ldots\ldots\ldots\ldots.(34.2)$$

وباستخدام المبدلة لـ e .

$$e'e = (Y - X\hat{B})'\left(Y - X\hat{B}\right)$$
$$e'e = \left(Y' - \hat{B}'X'\right)\left(Y - X\hat{B}\right)$$
$$e'e = Y'Y - \hat{B}'X'Y - Y'X\hat{B} + \hat{B}'X'X\hat{B} \ldots\ldots..(35.2)$$

بما أن الحد الثاني والثالث يمثلان قيمة واحدة وأن كل حد يمثل مبدلة للآخر فإن:

$$\left[\hat{B}'X'Y = \left(\hat{B}'X'Y\right)' = \hat{B}XY'\right]$$
$$\therefore e'e = Y'Y - 2\hat{B}'X'Y + \hat{B}'X'X\hat{B} \ldots\ldots\ldots..(36.2)$$

ولما كانت $\hat{B}' = \left(\hat{B}'\right)' = \hat{B}$

وبأخذ المشتقة الجزئية لـ $\hat{B}'$ ومساواتها بالصفر :

$$2X'X\hat{B} = 2X'Y$$

وبقسمة طرفي المعادلة على (2) نحصل:

$$X'X\hat{B} = X'Y$$

وبضرب طرفي المعادلة بالمعكوس $(X'X)^{-1}$ نحصل:

$$(X'X)^{-1}X'X\hat{B} = (X'X)^{-1}X'Y$$

وبما أن حاصل ضرب المعكوس $(X'X)^{-1}$ في المصفوفة $X'X$ يساوي مصفوفة الوحدة 1 ، إذًا المعادلة أعلاه تصبح:

$$\therefore \hat{B} = (X'X)^{-1}X'Y \dots\dots\dots (37.2)$$

حيث أن:

$\hat{B}$: تمثل معاملات الانحدار المطلوب تقديرها.

$(X'X)^{-1}$: تمثل معكوس المصفوفة $(X'X)$.

$X'Y$: تمثل المتجه.

وبالرجوع إلى البيانات الواردة في الجدول (1.2) وباعتماد صيغة المصفوفات يمكن الحصول على قيم $\hat{B}_0$ و $\hat{B}_1$ وكالاتي :

$$\hat{B} = (X'X)^{-1}X'Y$$

$$(X'X) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 4 & 8 & 12 & 16 & 20 & 24 & 28 & 32 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 1 & 8 \\ 1 & 12 \\ 1 & 16 \\ 1 & 20 \\ 1 & 24 \\ 1 & 28 \\ 1 & 32 \end{bmatrix}$$

$$(X'X) = \begin{bmatrix} 8 & 144 \\ 144 & 3264 \end{bmatrix}$$

$$(X'X)^{-1} = \frac{1}{|X'X|} adjX'X$$

$$|X'X| = \begin{bmatrix} 8 & 144 \\ 144 & 3264 \end{bmatrix} = (8)(3264) - (144)(144)$$

$$|X'X| = 26112 - 20736$$

$$|X'X| = 5376$$

$$adjX'X = \begin{bmatrix} 3264 & -144 \\ -144 & 8 \end{bmatrix}$$

$$(X'X)^{-1} = \frac{1}{5376} \begin{bmatrix} 3264 & -144 \\ -144 & 8 \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} 0.607142857 & -0.026785714 \\ -0.026785714 & 0.001488095 \end{bmatrix}$$

$$X'X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 4 & 8 & 12 & 16 & 20 & 24 & 28 & 32 \end{bmatrix} \begin{bmatrix} 25.6 \\ 32.7 \\ 45.4 \\ 53.9 \\ 59.0 \\ 62.6 \\ 65.0 \\ 65.8 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 410 \\ 8379.2 \end{bmatrix}$$

$$\hat{B} = (X'X)^{-1}X'Y$$

$$\hat{B} = \begin{bmatrix} 0.607142857 & -0.026785714 \\ -0.026785714 & 0.001488095 \end{bmatrix} \begin{bmatrix} 410 \\ 8379.2 \end{bmatrix}$$

$$\begin{bmatrix} \hat{B}_0 \\ \hat{B}_1 \end{bmatrix} = \begin{bmatrix} 248.9285714 & +(-224.4428547) \\ -10.98214274 & +12.46904562 \end{bmatrix}$$

$$\begin{bmatrix} \hat{B}_0 \\ \hat{B}_1 \end{bmatrix} = \begin{bmatrix} 24.48571 \\ 1.48690 \end{bmatrix}$$

$$\therefore\ \hat{Y}_i = 24.485 + 1.486\, X_i$$

يتضح مما سبق بأن النتائج التي تم التوصل إليها في المعادلة التقديرية هي نفسها تمامًا بالطرق الأربعة الآنفة الذكر.

3-2 الخواص الأساسية لمقدرات المربعات الصغرى:

في كل تقدير يتم الحصول عليه، هناك خصائص عدة مرغوب فيها لذلك التقدير، ومن هذه الخصائص خاصية أفضل مقدر خطي غير متحيز، Best Linear Unbiased Estimator (BLUE) فكل مقدر $(\widehat{\mathrm{B}})$ يمكن التعبير عنه كدالة خطية بالنسبة لمشاهدات المتغير التابع (Y)، أي أن:

$$\widehat{\mathrm{B}} = K_1Y_1 + K_2Y_2 + \cdots . + K_nY_n$$

حيث أن: K_i، $(\mathrm{i} = 1,2,\ldots,\mathrm{n})$ عبارة عن أوزان أو قيم ثابتة. ومن بين جميع المقدرات الخملية نبحث عن المقدرات غير المتحيزة، ونقصد بعدم التحيز هو أن يكون الفرق بين القيمة المتوقعة للمقدر وقيمة المعلمة الحقيقية يساوي صفر، أي: $I:(B) = B - 0$

وأفضل مقدر هو ذلك المقدر الذي يكون تباينه حول الوسط الحسابي أقل ما يمكن فإذا كان B^*، $\hat{B}$ مقدرات خطية غير متحيزة فإن $\widehat{\mathrm{B}}$ أفضل مقدر إذا تحققت العلاقة الآتية:

$$var(\hat{B}) < var\,(B^*)$$

وسوف نتناول أدناه هذه الخصائص بشيء من التفصيل:

1.2.3 الخاصية الخطية Linearity Property:

مقدرات المربعات الصغرى خطية في المتغير التابع حيث نلاحظ أن تلك المقدرات يمكن وصفها في صورة دالة أو ترتيب خطي من قيم المتغير Y، أي:

$$\hat{B}_1 = K_1Y_1 + K_2Y_2 + \cdots + K_nY_n$$

$$\therefore \hat{B}_1 = \sum_{i=1}^{n} K_iY_i$$

$$\therefore \hat{B}_1 = \frac{\sum x_i y_i}{\sum x_i^2}$$

وأن: للمشاهدة الواحدة: $y_i = Y_i - \bar{Y}$

$$\therefore \hat{B}_1 = \frac{\sum x_i\,(Y_i - \bar{Y})}{\sum x_i^2} = \frac{\sum x_i Y_i}{\sum x_i^2} - \frac{\bar{Y}\sum x_i}{\sum x_i^2}$$

$$\therefore \sum x_i = \sum (X_i - \bar{X}) = 0$$

$$\therefore \frac{\bar{Y}\sum x_i}{\sum x_i^2} = 0$$

لذلك تكون: $\hat{B}_1 = \frac{\sum x_i y_i}{\sum x_i^2} = \sum \left[\frac{x_i}{\sum x_i^2}\ Y_i\right]$

ولما كانت قيمX ثابتة نجد أن: $\frac{\sum x_i}{\sum x_i^2}$ هي مقادير ثابتة، ويمكن أن نرمز لها بالرمز K_i، وهي ثابتة في كل العينات، أي: $K_i = \frac{x_i}{\sum x_i^2}$

$\hat{B}_1$ مجموع مرجح لقيم المتغير التابع، Y_i

$$\therefore \hat{B}_1 = \sum K_iY_i$$

$$\therefore \hat{B}_1 = F(Y_i)$$

إذًا $\hat{B}_1$ هو مقدر خطي.

خصائص الأوزان Weighted Properties :

وبما أن الأوزان تعتمد على قيم X الثابتة فقط فإنها تعتبر ثابتة أيضًا، وتخضع K_i للشروط الآتية:

1-مجموع الأوزان يساوي صفر، أي أن:

$$\sum_{i=1}^{n} K_i = 0$$

إذًا:

$$\therefore K_i = \frac{x_i}{\sum x_i^2}$$

$$\therefore x_i = \sum (X_i - \bar{X}) = 0$$

وبالتعويض، فإن:

$$\therefore \sum K_i = \frac{0}{\sum x_i^2} = 0$$

2- مجموع حاصل الضرب للأوزان (K) في قيم المتغير المستقل (X_i) أو في انحرافاتها عن متوسطها الحسابي (x_i) يساوي الواحد صحيح، أي:

$$\sum_{i=1}^{n} K_i X_i = \sum_{i=1}^{n} k_i x_i = 1$$

$$\sum K_i x_i = \sum K_i (X_i - \bar{X}) = \sum K_i X_i - \bar{X} \sum K_i$$

وقد بينا سابقاً أن:

$$\because \sum K_i X_i = 0$$

$$\therefore \sum K_i x_i = \sum K_i X_i$$

وبما أن:

$$\because K_i = \frac{x_i}{\sum x_i^2}$$

إذًا:

$$\sum K_i x_i = \sum \frac{x_i}{\sum x_i^2} x_i$$

لذلك تكون:

$$\sum K_i x_i = \frac{\sum x_i^2}{\sum x_i^2} = 1$$

$$\therefore \sum_{i=1}^{n} K_i x_i = \sum_{i=1}^{n} K_i X_i = 1$$

3- مجموع مربعات الأوزان $\sum K_i^2$، يساوي معكوس مجموع مربعات انحرافات المتغير المستقل $\frac{1}{\sum x_i^2}$، أي أن:

$$K_i = \frac{x_i}{\sum x_i^2} = \frac{(X_i - X)}{\sum (X_i - \bar{X})^2}$$

وعند تربيع الطرفين، ثم جمعهما، يكون:

$$\sum K_i^2 = \frac{\sum(X_i - \bar{X})^2}{[\sum(X_i - \bar{X})^2]^2}$$

وبعد الاختصار والترتيب:

$$\sum K_i^2 = \frac{1}{\sum(X_i - \bar{X})^2} = \frac{1}{\sum x_i^2}$$

وبطريقة مماثلة يمكن البرهنة بأن $\hat{B}_0$ دالة خطية من قيم المتغير Y.

$$\hat{B}_0 = W_1Y_1 + W_2Y_2 \dots \dots . + W_nY_n$$

$$\therefore \hat{B}_0 = \sum_{i=1}^{n} W_iY_i$$

$$\therefore \hat{B}_0 = \bar{Y} - \hat{B}_1\bar{X}$$

وأن:

$$\bar{Y} = \frac{\sum Y_i}{n}, \hat{B}_1 = \sum K_iY_i$$

$$\therefore \hat{B}_0 = \frac{\sum Y_i}{n} - \bar{X}\sum K_iY_i$$

ومن خلال إعادة الترتيب، نحصل:

$$\hat{B}_0 = \sum \left[\frac{1}{n} - \bar{X}K_i\right] Y_i$$

ولما كانت K_i,$\bar{X}$ مقادير ثابتة، نرمز لها بالرمز W_i، أي أن:

$$\mathrm{W}_i = \left[\frac{1}{n} - \bar{X}K_i\right]$$

وبتعويض ذلك في المعادلة أعلاه، نجد أن $\hat{B}_0$ تعتمد على قيم Y_i فقط، بعبارة أخرى:

$$\hat{B}_0 = \sum W_oY_i$$

$$\therefore \hat{B}_0 = \mathrm{F}(Y_i)$$

إذًا $\hat{B}_0$ هو مقدر خطي.

2.2.3 عدم التحيز: (نظير مذكور، 2006–2007 صفحة 22)

تقصد بعدم التحيز هو أن يكون الفرق بين القيمة المتوقفة للمقدر $\mathrm{E}(\hat{\mathrm{B}})$ وقيمة المعلمة الحقيقية للمجتمع الإحصائي (B)، يساوي صفر، أي: $\mathrm{E}(\hat{\mathrm{B}}) - B = 0$

بعبارة أخرى يعتبر المقدر غير متحيز إذا كان وسطها يساوي القيمة الحقيقية للمعلمة: $\mathrm{E}(\hat{\mathrm{B}}) = B$

ولإثبات خاصية عدم التحيز بالنسبة $\hat{\mathrm{B}}_1$ نتبع ما يلي:

من خاصية الخطية: $\widehat{B}_1 = \sum K_i Y_i$

وباستحضار المعادلة: $Y_i = B_0 + B_1 X_i + U_i$

وبتعويض ذلك في أعلاه: $\widehat{B}_1 = \sum K_i \left(B_0 + B_1 X_i + U_i\right)$

وبفتح القوس نحصل:

$$\widehat{B}_1 = B_0 \sum K_i + B_1 \sum K_i X_i + \sum K_1 U_i \dots\dots\dots (38.2)$$

وباستخدام شروط الأوزان فإن:

$$\sum K_i = 0$$
$$B_0 \sum K_i = 0$$
$$\sum K_i X_i = 1$$

وبتعويض ذلك في المعادلة (38.2)، نحصل :

$$\therefore \hat{B}_1 = B_1 + \sum K_1 U_i \dots\dots\dots (39.2)$$

وبأخذ توقع طرفي المعادلة:

$$E(\hat{B}_1) = B_1 + \sum K_i \, E(U_i)$$

من فرضيات الخطأ العشوائي:

$$\because E(U_i) = 0$$
$$\because E K_i E(U_i) = 0$$
$$\therefore E(\hat{B}_1) = B_1 \dots\dots\dots\dots\dots\dots\dots (40.2)$$

ويعني ذلك أن $\hat{B}_1$ هو تقدير غير متحيز للقيمة الأصلية، B.

إثبات خاصية عدم التحيز لـ $\hat{B}_0$:

وبنفس المنهجية يمكن البرهنة بأن $\hat{B}_0$ تعتبر مقدرة غير متحيزة للمعلمة الحقيقية B_0.

من الخاصية الخطية يتبين لنا :

$$\hat{B}_0 = \sum W_i Y_i$$
$$W_i = \left[\frac{1}{n} - \bar{X} K_i\right]$$

وأن:

$$Y_i = B_0 + B_1 X_i + U_i$$

وعند التعويض نحصل:

$$\therefore \hat{B}_0 = \sum \left[\frac{1}{n} - \bar{X}K_i\right][B_0 + B_1X_i + U_i]$$

وعند فك الأقواس والتعويض، نحصل:

$$\hat{B}_0 = B_0 - B_0\bar{X}\sum K_i + B_1\bar{X} - \bar{X}B_1\sum K_iX_i + \sum\left(\frac{1}{n} - \bar{X}K_i\right)U_i \;\ldots\ldots (41.2)$$

وباستخدام شروط الأوزان:

$$\because \sum K_i = 0$$

$$\therefore B_0\bar{X}\sum K_i = 0$$

$$\because \sum K_iX_i = 1$$

$$\therefore \hat{B}_0 = B_0 + \sum\left(\frac{1}{n} - \bar{X}K_i\right)U_i \;\ldots\ldots. (42.2)$$

وبأخذ توقع طرفي المعادلة:

$$E(\hat{B}_0) = B_0 + \sum\left(\frac{1}{n} - \bar{X}K_i\right)E(U_i)$$

$$\because E(U_i) = 0$$

$$\therefore \sum\left(\frac{1}{n} - \bar{X}K_i\right)E(U_i) = 0$$

$$\therefore E(\hat{B}_0) = B_0 \;\ldots\ldots\ldots.. (43.2)$$

ويعني ذلك أن $\hat{B}_0$ هو تقدير غير متحيز للقيمة الأصلية، B_0

3.2.3.خاصية أفضل مقدر أقل تباين:

إن مفهوم تباين المعالم يُحدد بواسطة الانحراف بين المعالم المقدرة لـ B وقيمتها المتوقعة .$E(\hat{B})$

فبالنسبة لـ $\hat{B}_1$ فإن (نظير مدكور، 2006-2007 صفحة 28):

$$var(\hat{B}_1) = E\left[\hat{B}_1 - E(\hat{B}_1)\right]^2 \;\ldots\ldots\ldots\ldots (44.2)$$

من خاصية عدم التحيز : $E(\hat{B}_1) = B_1$

$$var(\hat{B}_1) = E\left[\hat{B}_1 - B_1\right]^2 \;\ldots\ldots\ldots\ldots (45.2)$$

بالرجوع إلى المعادلة (39.2) من خاصية عدم التحيز والخاصة بـ $\hat{B}_1$ نجد أن:

$$\hat{B}_1 = B_1 + \sum K_iU_i$$

$$\hat{B}_1 - B_1 = \sum K_iU_i$$

بتربيع الطرفين:

$$\left(\hat{B}_1 - B_1\right)^2 = \left(\sum K_i U_i\right)^2$$

وبأخذ توقع طرفي المعادلة:

$$E\left(\hat{B}_1 - B_1\right)^2 = E\left(\sum K_i U_i\right)^2$$

$$var\left(\hat{B}_1\right) = E\left[\sum K_i U_i\right]^2 \ldots\ldots\ldots\ldots\ldots (46.2)$$

وبفك القوس، نحصل:

$$var\left(\hat{B}_1\right) = \sum K_i^2\,(U_i)^2 + 2\sum_{i<j}\sum K_i U_j\, E\left(U_i U_j\right)$$

ومن فرضيات الخطأ العشوائي أن:

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

وأن: $E\left(U_i U_j\right) = 0 \qquad For\ all\ i \neq j$

$$\therefore var\left(\hat{B}_1\right) = \sum K_i^2\,\sigma^2 \ldots\ldots\ldots.(47.2)$$

$$\because K_i = \frac{x_i}{\sum x_i^2}$$

$$\sum K_i^2 = \frac{\sum x_i^2}{(\sum x_i^2)^2} = \frac{1}{\sum x_i^2}$$

$$\therefore var\left(\hat{B}_1\right) = \frac{1}{\sum x_i^2}.\sigma^2 = \frac{\sigma^2}{\sum x_i^2}$$

وبالأسلوب نفسه يمكن تحديد التباين لـ $\hat{B}_0$ على النحو الآتي:

$$\therefore var\left(\hat{B}_0\right) = E\left[\hat{B}_0 - E\left(\hat{B}_0\right)\right]^2 \ldots\ldots\ldots\ldots.(48.2)$$

من خاصية عدم التحيز : $E\left(\hat{B}_0\right) = B_0$

$$var\left(\hat{B}_0\right) = E\left[\hat{B}_0 - B_0\right]^2$$

وبالرجوع إلى المعادلة (42.2) من خاصية عدم التحيز والخاصة بـ $\hat{B}_0$ نجد أن:

$$\hat{B}_0 = B_0 + \sum\left(\frac{1}{n} - \bar{X}K_i\right)U_i$$

$$\hat{B}_0 - B_0 = \sum\left(\frac{1}{n} - \bar{X}K_i\right)U_i$$

$$var\left(\hat{B}_0\right) = E\left[\sum\left(\frac{1}{n} - \bar{X}K_i\right)U_i\right]^2$$

$$var\left(\hat{B}_0\right) = \sum\left(\frac{1}{n} - \bar{X}K_i\right)^2 E(U_i)^2$$

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

$$\therefore var(\hat{B}_0) = \sigma^2 \left[\sum \left(\frac{1}{n} - \bar{X} K_i \right)^2 \right]$$

وبفك القوس وباستخدام شروط الأوزان نحصل على:

$$\sum K_i = 0$$

$$\therefore \frac{2\bar{X}^2}{n} \sum K_i = 0$$

وأن:

$$\sum K_i^2 = \frac{1}{\sum X_i^2}$$

$$\therefore var(\hat{B}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2} \right)$$

ولإثبات أن مقدرات (OLS) هي أفضل مقدرات خطية غير متحيزة أي أنها تتسم بالكفاءة، ونتيجة لنسبية معيار الأفضلية فإننا نقوم بتعريف مقدرة أخرى ولتكن B_1^* تختلف عن مقدرة OLS ، $\hat{B}_1$ حيث سيتضح لاحقاً أن تباين تلك المقدرة B_1^* لابد أن يفوق تباين مقدرة OLS ، $\hat{B}_1$ وبالتالي نفضل المقدرة الأخيرة $\hat{B}_1$صاحبة التباين الأقل.

$$B_1^* = \sum C_i Y_i \ldots (49.2)$$

$$C_i = K_i + d_i$$

وأن $d_i \neq 0$ ، حيث d_i كميات ثابتة:

$$\therefore Y_i = B_0 + B_1 X_i + U_i$$

$$B_1^* = \sum C_i \left(B_0 + B_1 X_i + U_i \right) \ldots \ldots \ldots . (50.2)$$

$$B_1^* = B_0 \sum C_i + B_1 \sum C_i X_i + \sum C_i X U_i$$

ولكي تكون B_1^* غير متحيزة أي $E(B_1^*) = B_1$ يجب أن تتصف C_i بالصفات الآتية:

أ) $\sum C_i = 0$

أو: $\sum K_i + \sum d_i = 0$

معروفة مسبقًا : $\therefore \sum K_i = 0$

– $\sum d_i = 0$

أي إن: $\therefore \sum C_i = 0$

ب) $\sum C_i X_i = 1$

أو: $\because \sum K_i X_i + \sum d_i X_i = 1$

معروفة مسبقًا.

$$\sum K_i X_i = 1$$

أي أن: $\sum d_i X_i = 0$

$$\therefore \sum C\, X_i = 1 + 0 + 1$$

$$B_1^* = B_1 + \sum C_i U_i \dots\dots.. (51.2)$$

$$\because E(U_i) = 0$$

$$\therefore \sum C_i\, E(U_i) = 0$$

$$E(B_1^*) = B_1 \dots\dots\dots\dots (52.2)$$

هنا المقدر B_1^* غير متحيز ولتحديد تباين هذه المقدرة الخطية غير المتحيزة فإننا نعوض في قانون var.

$$varB_1^* = E[B_1^* - E(B_1^*)]^2$$

من المعادلة رقم (52.2): $\because\ E(B_1^*) = B_1$

$$varB_1^* = E[B_1^* - B_1]^2$$

وباستخدام المعادلة رقم (51.2):

$$B_1^* = B_1 + \sum C_i U_i$$

$$B_1^* - B_1 = \sum C_i U_i$$

$$.var(B_1^*) = E\left[\sum C_i U_i\right]^2 ..(53.2)$$

وبفك القوس، نحصل على:

$$var(B_1^*) = \sum C_i^2\, E(U_i)^2 + 2\sum_{i>j} C_i C_j\, E\big(C_i C_j\big)$$

ومن فرضيات الخطأ العشوائي: $E(U_i)^2 = \sigma^2$

وأن: $E\big(U_i U_j\big) = 0$

$$var(B_1^*) = \sum C_i^2\, \sigma^2 \dots\dots..(54.2)$$

ولما كانت: $C_i = K_i + d_i$

$$var(B_1^*) = \sum \Big(K_i + \sum d_i\Big)^2 \sigma^2$$

$$var(B_1^*) = \sigma^2 \left[\sum K_i^2 + \sum d_i^2 + \sum K_i d_i\right]$$

$$var(B_1^*) = \sigma^2 \sum K_i^2 + \sigma^2 \sum d_i^2 + 2\sigma^2 \sum K_i d_i$$

$$\because \sum K_i = 0$$

$$\because \sum K_i d_i = 0$$

$$var(B_1^*) = \sigma^2 \sum K_i^2 + \sigma^2 \sum d_i^2 \ldots\ldots\ldots.(55.2)$$

ولما كانت: $\sum K_i^2 = \frac{1}{\sum x_i^2}$

$$var(B_1^*) = \sigma^2 \frac{1}{\sum x_i^2} + \sigma^2 \sum d_i^2 \ldots\ldots\ldots\ldots(56.2)$$

$$\sigma^2 \frac{1}{\sum x_i^2} = var(B_1^*)$$

$$\therefore var(B_1) = var(\hat{B}_1) + \sigma^2 \sum d_i^2 \ldots\ldots\ldots..(57.2)$$

وحيث أن: $\sum d_i^2 > 0$

عليه فإن: $var(B_1^*) > var(\hat{B}_1)$

ومن ثم فإننا نرى ان مقدرة OLS الخاصة بالمعلمة $\hat{B}_1$ تتميز بأصغر تباين من بين المقدرات الخطية غير المتحيزة للمعلمة B_1^* .

4. تقدير تباين حد الخطأ العشوائي:

يستعمل تباين $\hat{B}_0$ و $\hat{B}_1$ في إجراء الاختبارات المعنوية الخاصة بتلك المقدرات، غير أن تباين المقدرات يحتوي على معلمة مجهولة هي (σ^2) تباين حد الخطأ العشوائي الذي يرمز له بالرمز (S_e^2) و يمكن اشتقاقه كالآتي:

بما أن المعادلة التقديرية تأخذ الصيغة الآتية:

$$\hat{Y}_i = \hat{B}_0 + \hat{B}_1 X_i \ldots\ldots..(58.2)$$

وبإدخال Σ على طرفي المعادلة :

$$\sum \hat{Y}_i = n\hat{B}_0 + \hat{B}_1 \sum X_i$$

وبالقسمة على n:

$$\frac{\sum \hat{Y}_i}{n} = \frac{n}{n}\hat{B}_0 + \hat{B}_1 \frac{\sum X_i}{n}$$

$$\overline{\hat{Y}} = \hat{B}_0 + \hat{B}_1 \bar{X} \ldots\ldots\ldots\ldots\ldots(59.2)$$

وبطرح المعادلة (59.2) من المعادلة (58.2) نحصل:

$$\hat{Y}_i - \overline{\hat{Y}} = \hat{B}_0 + \hat{B}_1 X_i - \hat{B}_0 - \hat{B}_1 \bar{X}$$

وبعد الاختصار، نحصل: $\hat{Y}_i = \hat{B}_1 X_i - \hat{B}_1 \bar{X}$

إذًا:

$$\hat{Y}_i = \hat{B}_1(X_i - \bar{X})$$

$$\hat{Y}_i = \hat{B}_1 x_i \ldots\ldots\ldots (60.2)$$

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

وبالتعويض من المعادلة (60.2):

$$\therefore e_i = y_i - \hat{B}_1 x_i$$

وبتربيع طرفي هذه المعادلة:

$$e_i^2 = \left(y_i - \hat{B}_1 x_i\right)^2$$

وعند فك الأقواس، نحصل:

$$e_i^2 = y_i^2 + \hat{B}_1^2 x_i^2 - 2\hat{B}_1 x_i y_i$$

وبإدخال Σ على طرفي المعادلة:

$$\sum e_i^2 = \sum y_i^2 + \hat{B}_1^2 \sum x_i^2 - 2\hat{B}_1 \sum x_i y_i \ldots\ldots (61.2)$$

$$\because \hat{B}_1 = \frac{\sum x_i y_i}{\sum x_i^2}$$

وإن حاصل ضرب الطرفين في الوسطين لذلك سيكون:

$$\hat{B}_1 \sum x_i^2 = \sum x_i y_i$$

وبالتعويض في المعادلة (61.2) بما يساويها:

$$\sum e_i^2 = \sum y_i^2 + \hat{B}_1 \sum x_i y_i - 2\hat{B}_1 \sum x_i y_i$$

وبعد الاختصار نحصل:

$$\sum e_i^2 = \sum y_i^2 + \hat{B}_1 \sum x_i y_i \ldots\ldots\ldots\ldots\ldots (62.2)$$

وقد جرى تعريف σ^2 على أنها: $\sigma^2 = \frac{\sum e_i^2}{n-2}$

وبالتعويض عن البسط من المعادلة 62.2 ، نحصل:

$$\sigma^2 = \frac{\sum y_i^2 - \hat{B}_1 \sum x_i y_i}{n-2} \ldots\ldots (63.2)$$

وهذا يعني أن تباين أخطاء العينة تمثل النسبة بين مجموع مربعات انحرافات القيم المشاهدة عن خط انحدار العينة إلى درجة الحرية للخطأ. وإذا رمزنا إلى التقدير الخطي غير المتحيز لتباين الخطأ بالرمز (S_e^2)، فإن:

$$S_e^2 = \frac{\sum e_i^2}{n-2}$$

ذلك يعني أن S_e^2 أفضل مقدر غير متحيز لتباين المتغير العشوائي أوحد الخطأ، U_i.

المحور الثالث:

تحليل الانحدار الخطي المتعدد

(خطوات صياغة نموذج متعدّد

تقدير معلمات النموذج، دراسة صلاحية

النموذج)

تمهيـد:

يتضح مما سبق أن الانحدار الخطي البسيط يركز على دراسة العلاقة بين متغيرين أحدهما المتغير المستقل (X) والآخر المتغير التابع (Y)، غير أن واقع الحياة الاقتصادية والاجتماعية مبني بشكل عام على تأثر أية ظاهرة بأكثر من متغير مستقل، فدالة الطلب مثلا، تحدد العلاقة بين الكميات المطلوبة من سلعة معينة، والاسعار الخاصة بهذه السلعة، وغيرها من السلع (أسعار السلع البديلة) ويدخل فيها دخل المستهلك كأحد المتغيرات فضلا عن المتغيرات الأخرى، لذلك لابد من توسيع نموذج الانحدار السابق ليشتمل على انحدار للمتغير التابع (Y) على العديد من المتغيرات المستقلة $(X_1, X_2, \ldots, X_k)$، ويسمى هذا بنموذج الانحدار الخطي المتعدِّد (Multiple Linear regression)

ويهدف هذا المحور إلى توضيح كيفية تقدير نموذج الانحدار الخطي المتعدِّد، الذي يتكون من متغير تابع ومتغيرين أو أكثر من المتغيرات المستقلة، لذلك سيتم مناقشة طبيعة نموذج الانحدار المتعدِّد، ثم تحديد أهم افتراضات النموذج التي سبق وأن تعرفنا عليها في المور السابق، يضاف إلى ذلك بيان عدم وجود علاقة خطية تامة بين المتغيرات المستقلة وكيف ان المصفوفة $(X'X)$، تكون مصفوفة غير شاذة $(Non-Singular)$ إذا كان محددها لا يساوي صفرا، ثم يتم بعد ذلك تقدير معلمات النموذج، تقدير التباين والتباين المشترك والانحراف المعياري لها للوصول إلى اختبار معاملات النموذج.

1.النموذج الخطي المتعدِّد:

يستند النموذج الخطي المتعدِّد على افتراض وجود علاقة بين متغير مستقل تابع

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \mu_i \ldots\ldots\ldots (I)$$

وفي واقع الأمر، فإن هذه المعادلة هي واحدة من جملة معادلات يبلغ عددها (n) تكون نظام المعادلات الآتي:

$$\begin{aligned}
Y_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \cdots + \beta_k X_{1k} + \mu_i \\
Y_i &= \beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \cdots + \beta_k X_{2k} + \mu_2 \\
&\vdots \\
Y_i &= \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \cdots + \beta_k X_{nk} + \mu_n
\end{aligned}$$

وهذه المعادلة تتضمن $(k+1)$ من المعلمات المطلوب تقديرها، علما أن الحد الأول منها (β_0) يمثل الحد الثابت، الأمر الذي يتطلب اللجوء إلى المصفوفات والمتجهات لتقدير تلك المعلمات (بخيت، وآخرون، 2006 الصفحات 145-144)، وعليه؛ يمكن صياغة هذه المعادلات في صورة مصفوفات كالتالي:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_3 \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} & X_{1k} \\ 1 & X_{21} & X_{22} & X_{23} & X_{2k} \\ 1 & X_{31} & X_{32} & X_{33} & X_{3k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{N1} & X_{n2} & X_{n3} & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \mu_0 \\ \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix}$$

وباختصار نقول:

$$Y = X\beta + \mu \dots\dots\dots (II)$$

Y: متجه عمودي أبعاده $(n \times 1)$ يحتوي مشاهدات المتغير التابع.

X: مصفوفة أبعادها $(n \times k + 1)$ تحتوي مشاهدات المتغيرات المستقلة، ويحتوي عمودها الأول على قيم الواحد الصحيح ليمثل الثابت.

β: متجه عمودي أبعاده $(k + 1 \times 1)$ ، ويحتوي على المعالم المطلوب تقديرها.

μ: متجه عمودي أبعاده $(n \times 1)$ يحتوي على الأخطاء العشوائية.

وبما أن المعادلة (II) هي العلاقة الحقيقية المجهولة والمراد تقديرها بإستخدام الإحصاءات المتوفرة عن المتغير التابع Y، والمتغيرات المستقلة $X_1, X_2, \dots, X_k$، فإنه يستوجب تحقق الفروض الأساسية الخاصة بـ μ_i التالية:

$$\mu_i \sim N(0, \sigma^2 In)$$

والذي يعني أن μ_i يتوزع توزيعا طبيعيا (N) متعدِّد المتغيرات لمتجه وسطه صفري (0) ومصفوفة تباين وتبيان مشترك عددية هي: $(\sigma^2 In)$.

1.1 فرضيات النموذج الخطي المتعدِّد:

عند استخدام طريقة (OLS) في تقدير نموذج الانحدار الخطي المتعدِّد/ فإنخ يجب توفر الافتراضات الآتية:

1. القيمة المتوقعة لمتجه حد الخطأ تساوي صفرا أي أن $E(\mu_i) = 0$

$$E(\mu_i) = E \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \vdots \\ \mu_n \end{bmatrix} = \begin{bmatrix} E(\mu_1) \\ E(\mu_2) \\ E(\mu_3) \\ \vdots \\ E(\mu_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0$$

- تباين العناصر العشوائية ثابت، والتباين المشترك بينها يساوي صفرا؛ أي أن:

$$Cov(\mu) = E(\mu\mu') = \sigma^2 In$$

$$E(\mu\mu') = E\begin{bmatrix}\mu_1\\ \mu_2\\ \vdots\\ \mu_n\end{bmatrix}[\mu_1 \quad \mu_2 \quad \cdots \quad \mu_n]$$

$$E(\mu\mu') = \begin{bmatrix}\mu_1^2 & \mu_1\mu_2 & \cdots & \mu_1\mu_n\\ \mu_2\mu_1 & \mu_2^2 & \cdots & \mu_2\mu_n\\ \vdots & \vdots & & \vdots\\ \mu_n\mu_1 & \mu_n\mu_2 & \cdots & \mu_n^2\end{bmatrix}$$

$$E(\mu\mu') = \begin{bmatrix}E(\mu_1^2) & E(\mu_1\mu_2) & \dots & E(\mu_1\mu_n)\\ E(\mu_2\mu_1) & E(\mu_2^2) & \dots & E(\mu_2\mu_n)\\ \vdots & \vdots & & \vdots\\ E(\mu_n\mu_1) & E(\mu_n\mu_2) & \cdots & E(\mu_n^2)\end{bmatrix}$$

وبما أن: $Var(\mu_i) = E(\mu_i^2) = \sigma^2$، فإن:

$$E(\mu\mu') = \begin{bmatrix}Var(\mu_1) & Cov(\mu_1\mu_2) & \dots & Cov(\mu_1\mu_n)\\ Cov(\mu_2\mu_1) & Var(\mu_2) & \dots & Cov(\mu_2\mu_n)\\ \vdots & \vdots & & \vdots\\ Cov(\mu_n\mu_1) & Cov(\mu_n\mu_2) & \cdots & Var(\mu_n)\end{bmatrix}$$

و بما أن $Cov(\mu_i\mu_j) = E(\mu_i\mu_j) = 0$ حيث أن $i \neq j$؛ يصبح لدينا:

$$E(\mu\mu') = \begin{bmatrix}\sigma_1^2 & 0 & \dots & 0\\ 0 & \sigma_2^2 & \dots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \sigma_n^2\end{bmatrix}$$

حيث أن: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2$، نجد ان:

$$E(\mu\mu') = \sigma^2\begin{bmatrix}1 & 0 & \dots & 0\\ 0 & 1 & \dots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & 1\end{bmatrix}$$

$$E(\mu\mu') = \sigma^2 In$$

وتسمى المصفوفة العددية أعلاه بمصفوفة التباين والتباين المشترك، $(Variance -$ $Covariance\ Matrix)$ لحد الخطأ (μ)، حيث تشكل العناصر القطرية في المصفوفة، تبيان قيم (μ) بينما تبقى العناصر غير القطرية (أعلى وأسفل القطر) (بخيت، وآخرون، 2006 صفحة 150)، مساوية للصفر لإنعدام التباين المشترك والترابط بين قين قيم (μ_i)

- ليس هناك علاقة خطية تامة بين المتغيرات المستقلة كما وأن عدد المشاهدات يجب أن يزيد على عدد المعلمات المطلوب تقديرها، أي أن:

$$r(x) = k + 1 < n$$

حيث أن (r) رتبة مصفوفة البيانات، (X) تساوي عدد المتغيرات المستقلة، (k) زائد (1) الحد الثابت، وهي أصغر من عدد المشاهدات (n)، وهذه الفرضية ضرورية جدا لضمان إيجاد معكوس المصفوفة $(X'X)$، إذ ان وجود هذا الفرض يجعل رتبة المصفوفة (X) أقل من $(k+1)$، وبالتالي فإن رتبة $(X'X)$ التي تستخدم في الحصول على مقدرات (OLS) بدورها أقل من $(k+1)$، ولا يمكن إيجاد معكوسها بسبب ما يسمى بمشكلة الارتباط الخطي المتعدِّد، وبالتالي لا يمكن الحصول على مقدرات المربعات الصغرى العادية (OLS).

- طرق تقدير معلمات النموذج:

في ضوء الفرضيات المذكورة أعلاه، يمكن استخدام (OLS)، في تقدير نعلمات النموذج الخطي المتعدِّد، ولهذا الغرض يمكن إعادة كتابة المعادلة (I) بصيغتاه التقديرية كالتالي:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2}$$

وبما أن هدفنا هو الحصول على قيم كل من $\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\right)$ التي تجعل مجموع مربعات الانحرافات أقل ما يمكن، أي تصغير القيمة $\left(\sum e_i^2\right)$ (مبدأ المربعات الصغرى) إلى أقل قيمة ممكنة، أي:

$$\min \rightarrow \sum_{i=1}^{n} e_i^2$$

$$e_i = Y_i - \hat{Y}_i$$

$$\sum e_i^2 = \sum \left(Y_i - \hat{Y}_i\right)^2$$

ومن خلال التعويض بـ $\hat{Y}_i$ بما يساويها وأخذ المشتقات الجزئية بالنسبة إلى $\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\right)$ ومساوتها بالصفر نحصل على:

$$\sum e_i^2 = \sum \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2}\right)^2$$

$$\frac{\partial \sum e_i^2}{\partial \hat{\beta}_0} = 2 \sum \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2}\right)(-1) = 0$$

$$2 \sum \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2}\right)(-1) = 0$$

بالقسمة على (-2) وفك القوس؛ نجد:

$$\sum Y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum X_{i1} - \hat{\beta}_2 \sum X_{i2} = 0$$

$$\sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_{i1} + \hat{\beta}_2 \sum X_{i2} \ldots\ldots\ldots (III)$$

$$\frac{\partial \sum e_i^2}{\partial \hat{\beta}_1} = 2 \sum \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2}\right)(-X_{i1}) = 0$$

$$-2 \sum X_{i1}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2}\right) = 0$$

بالقسمة على (-2) وفك القوس؛ نجد:

$$\sum X_{i1} Y_i - \hat{\beta}_0 \sum X_{i1} - \hat{\beta}_1 \sum X_{i1}^2 - \hat{\beta}_2 \sum X_{i1} X_{i2} = 0$$

$$\sum X_{i1} Y_i = \hat{\beta}_0 \sum X_{i1} + \hat{\beta}_1 \sum X_{i1}^2 + \hat{\beta}_2 \sum X_{i1} X_{i2} \dots\dots\dots (IV)$$

$$\frac{\partial \sum e_i^2}{\partial \hat{\beta}_2} = 2 \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2})(-X_{i2}) = 0$$

$$-2 \sum X_{i2} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2}\right) = 0$$

بالقسمة على (-2) وفك القوس؛ نجد:

$$\sum X_{i2} Y_i - \hat{\beta}_0 \sum X_{i2} - \hat{\beta}_1 \sum X_{i1} X_{i2} - \hat{\beta}_2 \sum X_{i2}^2 = 0$$

$$\sum X_{i2} Y_i = \hat{\beta}_0 \sum X_{i2} + \hat{\beta}_1 \sum X_{i1} X_{i2} + \hat{\beta}_2 \sum X_{i2}^2 \dots\dots\dots (V)$$

تمثل المعادلات (III) و (IV) و (V) المعادلات الطبيعية الثلاثة التي تستخدم في تقدير المعالم الثلاثة المجهولة $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$، ويمكن حل هذه المعادلات بالطرق التالية:

- طريقة المحددات:

ويمكن أن تحل هذه المعادلات بواسطة قاعدة كرايمر للحصول على قيم $\hat{\beta}_k$ من المعلمات وعلى النحو الآتي:

$$\sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_{i1} + \hat{\beta}_2 \sum X_{i2}$$

$$\sum X_{i1} Y_i = \hat{\beta}_0 \sum X_{i1} + \hat{\beta}_1 \sum X_{i1}^2 + \hat{\beta}_2 \sum X_{i1} X_{i2}$$

$$\begin{bmatrix} \sum Y_i \\ \sum X_{i1} Y_i \\ \sum X_{i2} Y_i \end{bmatrix} = \begin{bmatrix} n & \sum X_{i1} & \sum X_{i2} \\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i1} X_{i2} \\ \sum X_{i2} & \sum X_{i1} X_{i2} & \sum X_{i2}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

ومن خلال ذلك، نستطيع إيجاد المحددات التالية:

$$|D| = \begin{vmatrix} \sum Y_i & \sum X_{i1} & \sum X_{i2} \\ \sum X_{i1} Y_i & \sum X_{i1}^2 & \sum X_{i1} X_{i2} \\ \sum X_{i2} Y_i & \sum X_{i1} X_{i2} & \sum X_{i2}^2 \end{vmatrix}$$

$$|N_1| = \begin{vmatrix} \sum Y_i & \sum Y_i & \sum X_{i2} \\ \sum X_{i1} & \sum X_{i1}Y_i & \sum X_{i1}X_{i2} \\ \sum X_{i2} & \sum X_{i2}\,Y_i & \sum X_{i2}^2 \end{vmatrix}$$

$$|N_2| = \begin{vmatrix} n & \sum X_{i1} & \sum Y_i \\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i1}Y_i \\ \sum X_{i2} & \sum X_{i1}X_{i2} & \sum X_{i2}\,Y_i \end{vmatrix}$$

$$\hat{\beta}_1 = \frac{|N_1|}{|D|} = \frac{\begin{vmatrix} \sum Y_i & \sum Y_i & \sum X_{i2} \\ \sum X_{i1} & \sum X_{i1}Y_i & \sum X_{i1}X_{i2} \\ \sum X_{i2} & \sum X_{i2}\,Y_i & \sum X_{i2}^2 \end{vmatrix}}{\begin{vmatrix} \sum Y_i & \sum X_{i1} & \sum X_{i2} \\ \sum X_{i1}Y_i & \sum X_{i1}^2 & \sum X_{i1}X_{i2} \\ \sum X_{i2}\,Y_i & \sum X_{i1}X_{i2} & \sum X_{i2}^2 \end{vmatrix}}$$

$$\hat{\beta}_2 = \frac{|N_2|}{|D|} = \frac{\begin{vmatrix} n & \sum X_{i1} & \sum Y_i \\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i1}Y_i \\ \sum X_{i2} & \sum X_{i1}X_{i2} & \sum X_{i2}\,Y_i \end{vmatrix}}{\begin{vmatrix} \sum Y_i & \sum X_{i1} & \sum X_{i2} \\ \sum X_{i1}Y_i & \sum X_{i1}^2 & \sum X_{i1}X_{i2} \\ \sum X_{i2}\,Y_i & \sum X_{i1}X_{i2} & \sum X_{i2}^2 \end{vmatrix}}$$

أما بالنسبة لـ $\hat{\beta}_0$ فيتم الحصول عليه عن طريق:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}_1 - \hat{\beta}_2\bar{X}_2$$

- **طريقة الانحرافات:**

ويمكن تقدير معاملات الانحدار المتعدِّد باستخدام الانحرافات أو ما يسمى بالمتوسطات ، أي إنحرافات القيم الاصلية عن وسطها الحسابي كالآتي:

ولهذا نأخذ نموذج يحتوي متغيرين مستقلين X_1 و X_2:

$$\hat{Y}_i = \hat{\beta}_1 X_{i1} + \hat{\beta}_{i2} X_{i2} + e_i$$

وبأخذ المتوسط لهه المعادلة نجد:

$$\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1\bar{X}_1 + \hat{\beta}_2\bar{X} + \bar{e}_i \qquad , \qquad \bar{e}_i = 0$$

$$\hat{Y}_i - \bar{Y} = \hat{\beta}_1(X_{i1} - \bar{X}_1) + \hat{\beta}_2(X_{i2} - \bar{X}) + e_i$$

$$\hat{y}_i = \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + e_i \dots\dots\dots (V)$$

في واقع الأمر فإن المعادلة أعلاه هي واحدة من جملة معادلات يبلغ عددها n معادلة تكون نظام المعادلات التالي:

$$y_1 = \hat{\beta}_1 x_{11} + \hat{\beta}_2 x_{12} + \cdots + \hat{\beta}_k x_{1k} + e_1$$
$$y_2 = \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \hat{\beta}_k x_{2k} + e_2$$
$$\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots$$
$$y_n = \hat{\beta}_1 x_{n1} + \hat{\beta}_2 x_{n2} + \cdots + \hat{\beta}_k x_{nk} + e_n$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

حيث يمكن التعبير عن ذلك بصيغة المصفوفات كما يلي:

$$y = x\hat{\beta} + e$$

y: متجه عمودي أبعاده $(n \times 1)$ يحتوي على انحرافات قيم المتغير التابع.

x: مصفوفة أبعادها $(n \times k - 1)$ تحتوي على انحرافات قيم المتغيرات المستقلة، حيث أنها لا تتضمن العمود الأول الذي يمثل الحد لاثابت، حيث يمكن بذلك استخراج الحد الثابت $\hat{\beta}_0$ من خارج المصفوفة باستخدام القانون التالي:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$$

$\hat{\beta}$: متجه عمودي أبعاده $(k - 1 \times 1)$ تحتوي على المعالم المجهولة.

e: متجه عمودي ابعاده $(n \times 1)$ يحتوي على البواقي.

بإعادة كتابة المعادلة (V) على النحو التالي:

$$e_i = y_i - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}$$

ولما كانت أفضل طريقة للحصول على أصغر قيمة ممكنة للإنحرافات تتم بواسطة تربيعها وبجعل مربعاتها أصغر ما يمكن، وبأخذ المشتقة الجزئية لها بالنسبة لكل من $\hat{\beta}_1$ و $\hat{\beta}_2$ ومساواتها بالصفر؛ نحصل على ما يلي:

$$\sum e_i^2 = \sum \left(y_i - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}\right)^2$$

$$\frac{\partial \sum e_i^2}{\partial \hat{\beta}_1} = 2 \sum \left(y_i - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}\right)(-x_{i1}) = 0$$

$$-2 \sum x_{i1}\left(y_i - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}\right) = 0$$

وبالقسمة على $(2-)$ وفك القوس، نحصل على:

$$\sum x_{i1}\, y_i - \hat{\beta}_1 \sum (x_{i1})^2 - \hat{\beta}_2 \sum x_{i1}\, x_{i2} = 0$$

$$\sum x_{i1}\, y_i = \hat{\beta}_1 \sum (x_{i1})^2 + \hat{\beta}_2 \sum x_{i1}\, x_{i2} \dots \dots \dots (VI)$$

$$\frac{\partial \sum e_i^2}{\partial \hat{\beta}_2} = 2 \sum \left(y_i - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}\right)(-x_{i2}) = 0$$

$$-2 \sum x_{i2}\left(y_i - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2}\right) = 0$$

وبالقسمة على (-2) وفك القوس، نحصل على:

$$\sum x_{i2}\, y_i - \hat{\beta}_1 \sum x_{i1} x_{i2} - \hat{\beta}_2 \sum (x_{i2})^2 = 0$$

$$\dots \dots \dots (VII) \sum x_{i2}\, y_i = \hat{\beta}_1 \sum x_{i1} x_{i2} + \hat{\beta}_2 \sum (x_{i2})^2$$

ويمكن صياغة المعادلتين السابقين على شكل مصفوفة كما يلي:

$$\begin{bmatrix} \sum x_{i1}\, y_i \\ \sum x_{i2}\, y_i \end{bmatrix} = \begin{bmatrix} \sum (x_{i1})^2 & \sum x_{i1} x_{i2} \\ \sum x_{i1} x_{i2} & \sum (x_{i2})^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$$

ومن النظام أعلاه، يمكن إعادة كتابته بالشكل التالي:

$$x'y = (x'x)\hat{\beta}$$

وعليه فإن تقدير المعالم باستخدام المصفوفة بأسلوب الانحرافات يأخذ الصيغة التالية:

$$\hat{\beta} = x'y(x'x)^{-1}$$

وبعد احتساب المتجه $x'y$ ومحدد المصفوفة $|x'x|$ الذي ينبغي أن لا يساوي صفرا، نوجد مقلوب المصفوفة الذي هو عبارة عن:

$$(x'x)^{-1} = \frac{adj(x'x)}{|x'x|}$$

ومن ثم تطبيق القانون أعلاه، أما $\hat{\beta}_0$ فيمكن حسابه بموجب القانون التالي:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$$

هذا ويمكن استخراج القيم بالانحرافات دون الرجوع الى البيانات الاصلية كما يلي:

$$\sum x_1 y = \sum X_1 Y - \frac{(\sum X_1)(\sum Y)}{n}$$

$$\sum x_2 y = \sum X_2 Y - \frac{(\sum X_2)(\sum Y)}{n}$$

$$\sum x_1 x_2 = \sum X_1 X_2 - \frac{(\sum X_1)(\sum X_2)}{n}$$

$$\sum x_1^2 = \sum X_1^2 - \frac{(\sum X_1)^2}{n}$$

$$\sum x_2^2 = \sum X_2^2 - \frac{(\sum X_2)^2}{n}$$

$$\sum y_i^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n}$$

2. طرق تقدير نموذج الانحدار الخطي المتعدِّد:

لغرض التبسيط نفترض أن نموذج الانحدار الخطي المتعدِّد يتكون من متغير تابع ومتغيرين مستقلين (كامل علاوي، 2014 الصفحات 103-102). أي:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + U_i \dots\dots\dots\dots (4.9)$$

والعلاقة التقديرية للمعادلة (4.9) هي:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + e_i \dots\dots\dots (4.10)$$

وبتطبيق نظرية كاوس – ماركوف لتقدير معلمات النموذج:

$$e_i = Y_i - \hat{\beta}_0 X_{i1} - \hat{\beta}_2 X_{i2} \dots\dots\dots\dots (4.11)$$

إذ إن:

$$e_i = Y - \hat{Y}$$

وبتربيع طرفي المعادلة (4.11) وإدخال (Σ) عليها:

$$\sum e_i^2 = \sum \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2}\right)^2 \dots\dots (4.12)$$

ولتصغير مجموع مربعات الخطأ إلى أدنى ما يمكن، نأخذ المشتقات الجزئية للمعلمات في المعادلة (4.12) مع ($\sum e_i^2$) ومساواتها مع الصفر.

$$\frac{\partial \sum e_i^2}{\partial \hat{\beta}_0} = 2\sum\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2}\right)(-1) = 0$$

$$\sum Y_i = n\,\hat{\beta}_0 - \hat{\beta}_1 \sum X_{i1} - \hat{\beta}_2 \sum X_{i2} = 0$$

$$\sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_{i1} + \hat{\beta}_2 \sum X_{i2} \dots\dots\dots (4.13)$$

وبأخذ المشتقة الجزئية للعلاقة بين $\hat{\beta}_1$ و $\sum e_i^2$:

$$\frac{\partial \sum e_i^2}{\partial \hat{\beta}_1} = 2\sum\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2}\right)(-X_{i1}) = 0$$

$$\sum X_{i1} Y_i = \hat{\beta}_0 \sum X_{i1} + \hat{\beta}_1 \sum X_{i1}^2 + \hat{\beta}_2 \sum X_{i1} X_{i2} \dots\dots\dots (4.14)$$

وبنفس الطريقة نشتق العلاقة بين $\hat{\beta}_{i2}$ مع $\sum e_i^2$:

$$\sum X_{i1} Y_i = \hat{\beta}_0 \sum X_{i1} + \hat{\beta}_1 \sum X_{i1} X_{i2} + \hat{\beta}_2 \sum X_{i2}^2 \dots\dots\dots (4.15)$$

وبإعادة كتابة المعادلات (4.13) و (4.14) و (4.15):

$$\left.\begin{aligned}\sum Y_i &= n\hat{\beta}_0 + \hat{\beta}_1 \sum X_{i1} + \hat{\beta}_2 \sum X_{i2}\\ \sum X_{i1}Y_i &= \hat{\beta}_0 \sum X_{i1} + \hat{\beta}_1 \sum X_{i1}^2 + \hat{\beta}_2 \sum X_{i1}X_{i2}\\ \sum X_{i1}Y_i &= \hat{\beta}_0 \sum X_{i2} + \hat{\beta}_1 \sum X_{i1}X_{i2} + \hat{\beta}_2 \sum X_{i2}^2\end{aligned}\right\} \dots\dots..(4.16)$$

وتسمى منظومة المعادلات في (4.16) بالمعادلات الطبيعية، وتحل بعدة طرق مثل الحذف والتعويض وكرايمر والمصفوفات.

ويمكن كتابة المنظومة (4.16) على شكل مصفوفات وكالآتي: $Y = XB$

وباستخدام قانون المصفوفات الذي مر بنا في الانحدار البسيط المعادلة(3.50)[1] :

$$\hat{B} = (X'X)^{-1}X'Y \dots\dots\dots\dots\dots (4.17)$$

فإنّ:

$$(X'X) = \begin{bmatrix} 1 & 1 & \cdots & 1\\ X_{i1} & X_{i2} & \cdots & X_{in}\\ \vdots & \vdots & \ddots & \vdots\\ X_{n1} & X_{n2} & \cdots & X_{nn}\end{bmatrix}\begin{bmatrix} 1 & X_{11} & \cdots & X_{n1}\\ 1 & X_{12} & \cdots & X_{n2}\\ 1 & X_{11} & \cdots & X_{n3}\\ \vdots & \vdots & \ddots & \vdots\\ 1 & X_{1n} & \cdots & X_{nn}\end{bmatrix} \dots..(4.18)$$

$$= \begin{bmatrix} n & \sum X_{i1} & \sum X_{i2} & \cdots & \sum X_{in}\\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i1}X_{i2} & \cdots & \sum X\\ \vdots & \vdots & \vdots & & \vdots\\ \sum X_{in} & \sum X_{i1}X_{in} & \cdots & & \sum X_{in}^2\end{bmatrix}$$

وأن:

$$X'Y = \begin{bmatrix} 1 & 1 & \cdots & 1\\ X_{11} & X_{12} & \cdots & X_{1n}\\ \vdots & \vdots & \ddots & \vdots\\ X_{n1} & X_{2n} & \cdots & X_{nn}\end{bmatrix}\begin{bmatrix} Y_1\\ Y_2\\ Y_n\\ \vdots\\ Y_n\end{bmatrix} = \begin{bmatrix}\sum Y\\ \sum X_{i1}Y\\ \sum X_{i2}Y\\ \vdots\\ \sum X_{in}Y\end{bmatrix} \dots\dots (4.19)$$

[1] - إن المعادلة (4.17) صحيحة في حالة الانحدار البسيط أو المتعدد لذا تم الاقتصار على ذكرها فقط.

مثال11: استخدم البيانات الآتية لتقدير منظومة المعادلات الآنية الواردة في النموذج (4.16):

Y	X_1	X_2	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y
7	1	2	1	4	2	7	14
8	2	1	4	1	2	16	8
5	1	3	1	9	3	5	15
6	3	1	9	1	3	18	6
4	1	2	1	4	2	4	8
30	8	9	16	19	12	50	51

وبالتعويض في منظومة المعادلات (4.16):

$$30 = 5\hat{\beta}_0 + 8\hat{\beta}_1 + 9\hat{\beta}_2 \dots\dots..(1)$$
$$30 = 5\hat{\beta}_0 + 8\hat{\beta}_1 + 9\hat{\beta}_2 \dots\dots\dots..(1)$$
$$50 = 8\hat{\beta}_0 + 16\hat{\beta}_1 + 12\hat{\beta}_2 \dots\dots..(2)$$
$$51 = 9\hat{\beta}_0 + 12\hat{\beta}_1 + 19\hat{\beta}_2 \dots\dots..(3)$$

نضرب المعادلة (1) بالعدد الصحيح (2):

$$60 = 10\hat{\beta}_0 + 19\hat{\beta}_1 + 18\hat{\beta}_2 \dots\dots..(4)$$

وبطرح المعادلة (2) من المعادلة (4):

$$10 = 2\hat{\beta}_0 + 6\hat{\beta}_2 \dots\dots..(5)$$
$$\hat{\beta}_0 = 5 - 3\hat{\beta}_2 \dots\dots\dots..(6)$$

وبضرب المعادلة (2) بالعدد الصحيح (2)، والمعادلة (3) بالعدد الصحيح (4):

$$150 = 24\hat{\beta}_0 + 48\hat{\beta}_1 + 36\hat{\beta}_2 \dots\dots..(7)$$
$$204 = 36\beta_0 + 48\hat{\beta}_1 + 76\hat{\beta}_2 \dots\dots..(8)$$

وبالطرح:

$$54 = 12\beta_0 + 40\beta_2 \dots\dots..(9)$$
$$54 = 12(5 - 3\beta_2) + 40\beta_2 \dots\dots..(10)$$
$$-6 = 4\beta_2$$
$$\beta_2 = -1.5$$

وبالتعويض عن قيمة (β_2) في المعادلة (6) نحصل على:

$$\hat{\beta}_0 = 5 - 3(-1.5) = 9.5$$

أما قيمة (β_1):

$$30 = 5(9.5) + 8\hat{\beta}_1 + 9(-1.5) \dots\dots\dots\dots.(1)$$
$$30 = 47.5 + 8\hat{\beta}_1 - 13.5$$
$$-4 = 8\hat{\beta}_1$$

$$\hat{\beta}_1 = \frac{-4}{8} = -0.5$$

وبذلك تكون المعادلة التقديرية كالآتي:

$$\hat{Y} = 9.5 - 0.5X_1 - 1.5X_2$$

وبالاختصار على متغيرين مستقلين نستخدم المنظومة (4.16) في الحل.

أما تباين المعلمات فنحصل عليه باستخدام القانون الآتي:

$$var(\hat{\beta}_1) = \sigma_u^2 (X'X)^{-1} \dots\dots\dots.(4.20)$$

التقدير باستخدام طريقة المصفوفات (كامل علاوي، 2014 الصفحات 107-106):

$$\sum Y = 30 \qquad \sum X_1 = 8 \qquad \sum X_2 = 9 \qquad \sum X_1X_2 = 12$$

$$\sum X_1^2 = 16 \qquad \sum X_2^2 = 19 \qquad \sum X_1Y = 50 \qquad \sum X_2Y = 51$$

وبتطبيق القانون الآتي:

$$\hat{B} = (X'X)^{-1}X'Y \dots\dots\dots\dots\dots\dots (4.17)$$

$$X'X = \begin{bmatrix} n & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1^2 & \sum X_1X_2 \\ \sum X_2 & \sum X_2X_1 & \sum X_2^2 \end{bmatrix} = \begin{bmatrix} 5 & 8 & 9 \\ 8 & 16 & 12 \\ 9 & 12 & 19 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} \sum Y \\ \sum X_1Y \\ \sum X_2Y \end{bmatrix} = \begin{bmatrix} 30 \\ 50 \\ 51 \end{bmatrix}$$

$$(X'X)^{-1} = \frac{1}{|X'X|} adjX'X$$

$$|X'X| = 5\begin{vmatrix} 16 & 12 \\ 12 & 19 \end{vmatrix} - 8\begin{vmatrix} 8 & 12 \\ 9 & 19 \end{vmatrix} + 9\begin{vmatrix} 8 & 16 \\ 9 & 12 \end{vmatrix} = 16$$

$$(X'X)^{-1} = \frac{1}{16}\begin{bmatrix} 160 & -44 & -48 \\ -44 & 14 & 12 \\ -48 & 12 & 16 \end{bmatrix}$$

$$\hat{\beta}_1 = (X'X)^{-1}X'Y = \frac{1}{16}\begin{bmatrix} 160 & -44 & -48 \\ -44 & 14 & 12 \\ -48 & 12 & 16 \end{bmatrix} = \begin{bmatrix} 30 \\ 50 \\ 51 \end{bmatrix}$$

$$\hat{\beta}_0 = \frac{1}{16}[160 \cdot 30 - 44 \cdot 50 - 48 \cdot 51] = 9.5$$

$$\hat{\beta}_1 = \frac{1}{16}[-44 \cdot 30 + 14 \cdot 50 + 128 \cdot 51] = -1.5$$

$$\hat{\beta}_2 = \frac{1}{16}[-48 \cdot 30 + 12 \cdot 50 + 16 \cdot 51] = -0.5$$

وبذلك تكون المعادلة التقديرية كالآتي:

$$\hat{Y} = 9.5 - 0.5X_1 - 1.5X_2$$

3.التقدير حول نقطة المتوسط (استخدام الانحرافات):

يمكن استخدام انحرافات القيم عن أوساطها الحسابية في تقدير معالم الانحدار الخطي المتعدِّد. وبافتراض متغيرين مستقلين (X_1 و X_2) فإن المعادلة التقديرية هي:

$$Y_1 = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + e_i \dots\dots.(4.21)$$
$$\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}_1 + \hat{\beta}_2 \bar{X}_2 \dots\dots.(4.22)$$
$$\mathrm{y} = \hat{\beta}_1(X_{i1} + \bar{X}_1) + \hat{\beta}_2(X_{i2} + \bar{X}_2) + e_i$$
$$y = \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + e_i \dots\dots\dots\dots\dots.(4.23)$$
$$e_i = y - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} \dots\dots\dots.(4.24)$$

وباستخدام طريقة كاوس – ماركوف وكما مر سابقا:

$$\sum e_i^2 = \sum\left(-\hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2}\right)^2 \dots\dots\dots\dots(4.25)$$
$$\frac{\partial \sum e_i^2}{\partial \hat{\beta}_1} = 2\sum\left(y - \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2}\right)(-X_{i1}) = 0$$
$$\sum x_{i1} y_i = \hat{\beta}_1 \sum x_{i1}^2 - \hat{\beta}_2 \sum x_{i1} x_{i2} \dots\dots\dots\dots\dots(4.26)$$
$$\frac{\partial \sum e_i^2}{\partial \hat{\beta}_2} = 2\sum\left(y - \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2}\right)(-X_{i2}) = 0$$
$$\sum x_{i2} y_i = \hat{\beta}_1 \sum x_{i1} x_{i2} - \hat{\beta}_2 \sum x_{i2}^2 \dots\dots\dots\dots\dots(4.27)$$

أما الحد الثابت فقيد وفق الصيغة الرياضية الآتية:

$$\hat{\beta}_0 = \bar{Y} + \hat{\beta}_1 \bar{X}_1 + \hat{\beta}_2 \bar{X}_2 \dots\dots\dots\dots\dots\dots.(4.28)$$

والمعادلتان (4.26) و (4.27) هي المعادلات الطبيعية باستخدام البيانات بانحرافاتها عن أوساطها الحسابية وتحل بالطرق مارة الذكر.

$$\left.\begin{aligned}\sum x_{i1} y_i = \hat{\beta}_1 \sum x_{i1}^2 - \hat{\beta}_2 \sum x_{i1} x_{i2} \\ \sum x_{i2} y_i = \hat{\beta}_1 \sum x_{i1} x_{i2} - \hat{\beta}_2 \sum x_{i2}^2\end{aligned}\right\} \dots\dots\dots.(4.29)$$

وبكتابة المنظومة (4.29) على شكل مصفوفات:

$$\begin{bmatrix} \sum x_{i1}y_i \\ \sum x_{i2}y_i \end{bmatrix} = \begin{bmatrix} \sum x_{i1}^2 & \sum x_{i1}x_{i2} \\ \sum x_{i1}x_{i2} & \sum x_{i2}^2 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \dots\dots\dots (4.30)$$

وباستخدام قاعدة كرايمر نجد:

$$\hat{\beta}_1 = \frac{|A_1|}{|D|} = \frac{\begin{bmatrix} \sum x_{i1}y_i & \sum x_{i1}x_{i2} \\ \sum x_{i2}y_i & \sum x_{i2}^2 \end{bmatrix}}{\begin{bmatrix} \sum x_{i2}^2 & \sum x_{i1}x_{i2} \\ \sum x_{i1}x_{i2} & \sum x_{i2}^2 \end{bmatrix}}$$

$$\hat{\beta}_1 = \frac{\sum x_{i1}y_i \sum x_{i2}^2 - \sum x_{i2}y_i \sum x_{i1}x_{i2}}{\sum x_{i1}^2 \sum x_{i2}^2 - (\sum x_{i1}x_{i2})^2} \dots\dots\dots (4.31)$$

$$\hat{\beta}_2 = \frac{|A_2|}{|D|} = \frac{\begin{bmatrix} \sum x_{i2}^2 & \sum x_{1i}y_i \\ \sum x_{i1}x_{i2} & \sum x_{i2}y_i \end{bmatrix}}{\begin{bmatrix} \sum x_{i1}^2 & \sum x_{i1}x_{i2} \\ \sum x_{i1}x_{i2} & \sum x_{i2}^2 \end{bmatrix}}$$

$$\hat{\beta}_2 = \frac{\sum x_{i1}^2 \sum x_{i1}y_i - \sum x_{i2}y_i \sum x_{i1}x_{i2}}{\sum x_{i1}^2 \sum x_{i2}^2 - (\sum x_{i1}x_{i2})^2} \dots\dots\dots (4.32)$$

وباستخدام المصفوفات تقدر معالم النموذج وفق القانون الآتي:

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = (x'x)^{-1}x'y \dots\dots\dots\dots (4.33)$$

إذ أن:

$$(x'x) = \overbrace{\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \end{bmatrix}}^{x'} \overbrace{\begin{bmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \\ \vdots & \vdots \\ x_{1n} & x_{2n} \end{bmatrix}}^{x} = \begin{bmatrix} \sum x_{i1}^2 & \sum x_{11}x_{12} \\ \sum x_{i1}x_{i2} & \sum x_{i2}^2 \end{bmatrix} \dots\dots (4.34)$$

$$(x'y) = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum x_{i1}y \\ \sum x_{i2}y \end{bmatrix} \dots\dots\dots (4.35)$$

أما التباين والتباين المشترك فنحصل عليه من القانون الآتي:

$$var \begin{bmatrix} \hat{B}_1 \\ \hat{B}_2 \end{bmatrix} = \sigma_u^2 (x'x)^{-1} \dots\dots\dots\dots (4.36)$$

وتباين الحد الثابت:

$$var\hat{B}_0 = \sigma_u^2 \left[\bar{X}_i' (x'x)^{-1} \bar{X}_i + \frac{1}{n} \right] \dots\dots\dots\dots (4.37)$$

إذ إن: $\sigma_u^2 = \frac{\sum_{i=1}^{n} e^2}{N-K}$

مثال (12): البيانات الآتية تمثل العلاقة بين المتغير المعتمد (Y) والمتغيرات المستقلة (X_1) و (X_2) لعينة من خمسة مشاهدات:

Y	X_1	X_2	y	x_1	x_1^2	x_2	x_2^2	x_1y	x_2y	x_1x_2	y^2
40	4	8	−22	−4	16	−0.4	0.16	88	8.8	1.6	484
60	6	12	−2	−2	4	3.6	12.96	4	−7.2	−7.2	4
50	7	10	−12	−1	1	1.6	2.56	12	−19.2	−1.6	144
70	10	5	8	2	4	−3.4	11.56	16	−27.8	−6.8	64
90	13	7	28	5	25	−1.4	1.96	140	−39.2	−7	784
310	40	42	−−	−−	50	−−	29.7	260	−84	−21	1480

<u>الحل</u>: تقدير النموذج بطريقة المعادلات الطبيعية:

$$\left.\begin{aligned} \sum x_{i1}y_i &= \hat{\beta}_1 \sum x_{i1}^2 - \hat{\beta}_2 \sum x_{i1}x_{i2} \\ \sum x_{i2}y_i &= \hat{\beta}_1 \sum x_{i1}x_{i2} - \hat{\beta}_2 \sum x_{i2}^2 \end{aligned}\right\} \ldots\ldots\ldots.(4.29)$$

$$260 = 50\hat{\beta}_1 - 21\hat{\beta}_2 \ldots\ldots\ldots..\ldots..(1)$$

$$-84 = -21\hat{\beta}_1 - 29.7\hat{\beta}_2 \ldots\ldots\ldots..(2)$$

نضرب المعادلة (1) بـ (21) والمعادلة (2) بـ (50):

$$5460 = \cancel{1050}\hat{\beta}_1 - 441\hat{\beta}_2 \ldots\ldots\ldots..(3)$$

$$-4200 = -\cancel{1050}\hat{\beta}_1 - 1485\hat{\beta}_2 \ldots\ldots.(4)$$

بالجمع نحصل على:

$$1260 = 1044\hat{\beta}_2$$

$$\hat{\beta}_2 = 1.2069$$

$$260 = 50\hat{\beta}_1 - 21(1.2069)$$

$$285.3449 = 50\hat{\beta}_1$$

$$\hat{\beta}_1 = 5.707$$

$$\hat{\beta}_0 = \bar{Y} + \hat{\beta}_1\bar{X}_1 + \hat{\beta}_2\bar{X}_2$$

$$\hat{\beta}_0 = 62 - 5.707(8) - 1.2069(8.4) = 6.20604$$

المعادلة التقديرية هي:

$$\hat{Y} = 6.20604 + 5.707X_1 + 1.2064X_2$$

1) التقدير بطريقة كرايمر: نعيد كتابة المعادلات بصيغة المصفوفات:

$$\begin{bmatrix}\sum x_{i1}y_i\\ \sum x_{i2}y_i\end{bmatrix}=\begin{bmatrix}\sum x_{i1}^2 & \sum x_{i1}x_{i2}\\ \sum x_{i1}x_{i2} & \sum x_{i2}^2\end{bmatrix}=\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{bmatrix}\dots\dots\dots.(4.30)$$

$$\begin{bmatrix}260\\ -84\end{bmatrix}=\begin{bmatrix}50 & -21\\ -21 & 29.7\end{bmatrix}\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{bmatrix}\dots\dots\dots\dots.(5)$$

$$\hat{\beta}_1=\frac{|A_1|}{|D|}=\frac{\begin{bmatrix}260 & -21\\ -84 & 29.7\end{bmatrix}}{\begin{bmatrix}50 & -21\\ -21 & 29.7\end{bmatrix}}=\frac{5958}{1044}=5.707$$

$$\hat{\beta}_2=\frac{|A_2|}{|D|}=\frac{\begin{bmatrix}50 & 260\\ -21 & -84\end{bmatrix}}{\begin{bmatrix}50 & -21\\ -21 & 29.7\end{bmatrix}}=\frac{1260}{1044}=1.2069$$

أما الحد الثابت فيتم تقديره كما مر سابقا.

3) التقدير بطريقة معكوس المصفوفة:

$$\hat{B}_1=(x'x)^{-1}x'y$$

$$\hat{B}_1=\frac{1}{|x'x|}adj\ x'x\cdot x'y$$

من المصفوفة (5):

$$\hat{\beta}_1=\frac{1}{1044}\begin{bmatrix}29.7 & 21\\ 21 & 50\end{bmatrix}\begin{bmatrix}260\\ -84\end{bmatrix}$$

$$\hat{\beta}_1=\frac{1}{1044}[29.7\ \cdot 260-84\cdot 21]=5.707$$

$$\hat{\beta}_2=\frac{1}{1044}[21\ \cdot 260-50\cdot 84]=1.2069$$

أما الحد الثابت فيقدر بالطريقة السابقة.

4.تحليل الانحرافات في حالة الانحدار الخطي المتعدِّد:

لقد أوضحنا بأن الانحرافات الكلية $\sum_{i=1}^{n}y^2=\sum_{i=1}^{n}(Y-\bar{Y})^2$ تنقسم إلى جزأين، الأول يمثل الانحرافات الموضحة $\sum_{i=1}^{n}\bar{y}^2=\sum_{i=1}^{n}\left(\hat{Y}-\bar{Y}\right)^2$ والثاني الانحرافات غير الموضحة $\sum_{i=1}^{n}\hat{e}_i^2=\sum_{i=1}^{n}(Y-\bar{Y})^2$ أي إن:

$$\sum_{i=1}^{n}y^2=\sum_{i=1}^{n}\hat{y}^2+\sum_{i=1}^{n}\sigma_1^2\dots\dots\dots\dots(4.38)$$

$$SST=SSR+SSE$$

إذ إن:

الانحرافات الكلية: SST

الانحرافات الموضحة: SSR

الانحرافات غير الموضحة: SSE

وبما إن معامل التحديد المتعدِّد (R^2) يعد مؤشرًا على جودة توفيق خط الانحدار، فهو يحسب بقسمة الانحرافات الموضحة على الانحرافات الكلية.

$$R^2 = \frac{\sum_{i=1}^{n} \hat{y}^2}{\sum_{i=1}^{n} y^2} = \frac{SSR}{SST}$$

وإذا كان لدينا (K) من المتغيرات فإن:

$$R^2 = \frac{\hat{\beta}_1 \sum_{i=1}^{n} x_{i1} y + \hat{\beta}_2 \sum_{i=1}^{n} x_{i2} y + \cdots .. + \hat{\beta}_k \sum_{i=1}^{n} x_{ik} y}{\sum_{i=1}^{n} y_i^2} \dots \dots . (4.39)$$

وباستخدام المصفوفات يمكن أن نشتق صيغة معامل التحديد المتعدِّد (R^2):

$$e'e = \left(y - x\hat{\beta}\right)'\left(y - x\hat{\beta}\right)$$
$$e'e = y'y - y'x\hat{\beta} - \hat{\beta}'yx + \hat{\beta}'x'x\hat{\beta}$$
$$e'e = y'y - \hat{\beta}'x'y + \hat{\beta}'x'x\hat{\beta}$$
$$e'e = y'y - \hat{\beta}'x'y \dots \dots \dots . (4.40)$$

إذ إن:

الانحرافات غير الموضحة: $e'e$

الانحرافات الكلية $y'y$

الانحرافات الموضحة $\hat{\beta}'x'y$

وبإعادة كتابة المعادلة: $y'y = \hat{\beta}'x'y + e'e$

ومنها: $R^2 = \frac{\hat{\beta}' x' y}{y' y} \dots \dots \dots . (4.41)$

ويمكن الحصول على الانحرافات الموضحة وغير الموضحة بدلالة معامل التحديد المتعدِّد (R^2) وكالآتي:

$$\hat{\beta}'x'y = R^2 \ y'y \dots \dots \dots \dots . (4.42)$$
$$e'e = (1 - R^2)y'y \dots \dots \dots (4.43)$$

أما باستخدام البيانات الأصلية تحول إلى انحرافات للحصول على (R^2) وكالاتي:

$$R^2 = \frac{\hat{\beta}'X'Y - n\bar{Y}^2}{Y'Y - n\bar{Y}^2} \dots \dots . . (4.44)$$

إن إضافة متغيرات مستقلة إلى نموذج الانحدار الخطي المتعدِّد يجعل من معامل التحديد المتعدِّد (R^2) متحيز علويا Upward Based، إذ إن إضافة المتغيرات تؤدي إلى تخفيض درجات الحرية $(N-K)$ مما يتطلب تصحيح المعامل وفق الصيغة الآتية:

$$\bar{R}^2 = 1 - \left[(1-R^2)\frac{n-1}{n-k}\right] \dots\dots\dots\dots (4.45)$$

إذ إن معامل التحديد المصحح $(\bar{R}^2)$.

ومما تجدر الإشارة إليه إن قيمة معامل التحديد تتراوح بين الصفر والواحد: $0 \le R^2 \le 1$

في حين يمكن أن يكون معامل التحديد المصحح $(\bar{R}^2)$ سالبًا، إذا كان عدد المتغيرات كبيرًا وحجم العينة (n) صغيراً.

وبالرجوع إلى المثال (12) يمكن الحصول على (R^2):

$$\mathrm{R}^2 = \frac{\hat{\beta}_1 \sum_{i=1}^{n} x_{i1}y + \hat{\beta}_2 \sum_{i=1}^{n} x_{i2}y}{\sum_{i=1}^{n} y_i^2} = \frac{5.707(260) + 1.2069(-84)}{1480}$$

$$= \frac{1382.4404}{1480} = 0.93$$

$$\mathrm{R}^2 = \frac{\hat{\beta}'X'Y}{Y'Y} = \frac{[5.707 \quad 1.2069]\begin{bmatrix}260\\-84\end{bmatrix}}{1480} = \frac{1382.4404}{1480} = 0.93$$

وهذه تفسر أن 93% من التغيرات التي تحدث في المتغير التابع سببها المتغيرات المستقلة الداخلة في النموذج.

$$\bar{R}^2 = 1 - \left[(1-R^2)\frac{n-1}{n-k}\right] = 1 - \left[(1-0.93)\frac{5-1}{5-3}\right] = 0.86$$

5. اختبار الفروض Hypothesis Test:

يكتسب تحليل عملية اختبار الفروض أهمية في تحليل الانحدار الخطي المتعدِّد (عبد القادر عطية، 2004 صفحة 313)، وذلك كونها تعطي الباحث مؤشرًا على استبعاد المتغيرات المستقلة التي لا تمارس تأثيرها على المتغير التابع. لذا تنطوي فرضية العدم على عدم وجود العلاقة بين المتغيرات المستقلة والمتغير التابع، أي إن مصدر الانحرافات هو المتغير العشوائي فقط، وعادة ما يستخدم اختبار (F) في ذلك. وتصاغ فرضية العدم كالاتي:

$$H_0 = \hat{\beta}_0 = \hat{\beta}_1 = \cdots \dots\dots. = \hat{\beta}_k = 0 \dots\dots. (4.46)$$

أما إذا تم رفض الفرضية أعلاه أي هناك علاقة خطية بين المتغيرات المستقلة والمتغير التابع فإن الفرض البديل يصاغ بالشكل الآتي:

$$H_1 = \hat{\beta}_0 \neq \hat{\beta}_1 \neq \cdots \ldots \ldots \neq \hat{\beta}_k \neq 0 \ldots \ldots .. (4.47)$$

ويحسب اختبار (F) حسب الصيغة الآتية:

$$F = \frac{\hat{\beta}' x' y / k-1}{e' e / n-k} = \frac{y' y R^2 / k-1}{y' y\,(1-R^2)/n-k}$$

$$F = \frac{\sum \hat{y} 2 / k-1}{\sum e^2 / n-k}$$

$$F = \frac{R^2 / k-1}{(1-R^2)/n-k} \ldots \ldots . (4.48)$$

وتكوين قاعدة القرار تحدد كالآتي:

إذا كانت القيمة المحتسبة لاختبار (F) أكبر من القيمة الجدولية بمستوى معنوي معين، وبدرجات حرية (V_1) للبسط و (V_2) للمقام، نرفض فرضية العدم ونقبل الفرض البديل والعكس في حالة كون (F) المحتسبة أقل من الجدولية إذ نقبل فرضية العدم ونرفض الفرض البديل (عبد القادر عطية، 2004 صفحة 318).

ويمكن توضيح الانحرافات في جدول تحليل التباين وكالآتي:

Anova Table

Source of Varriation	**Sun of Square**	**d.f**	**Mean Error Sun**	**F**
Explained Variation	$\hat{\beta}' x' y = \sum \hat{y}^2 = y' y R^2$	$k-1$	$\frac{\sum \hat{y}^2}{k-1}$	$F = \frac{\sum \hat{y}^2}{k-1}$
Unexplained Variation	$\sum e^2 = y' y (1-R^2) = y' y - \hat{\beta}' x' y$	$N-K$	$\sum e^2 / N-K$	
Total Variation	$y' y = \sum y^2$	$n-1$		

وباستخدام البيانات السابقة يمكن بناء جدول تحليل التباين وكالاتي:

Anova Table

Source of Varriation	**Sun of Square**	**d.f**	**Mean Error Sun**	**F**
Explained Variation	$\sum \hat{y}^2 = 13824404$	$k-1 = 3-1 = 2$	$\frac{\sum \hat{y}^2}{k-1} = \frac{1382.4404}{2} = 691.2202$	$F = \frac{\sum \hat{y}^2 / k-1}{\sum e^2 / k-1} = \frac{691.2202}{48.7798} = 14.17$
Unexplained Variation	$\sum e^2 = 97.5596$	$N-K = 5-3 = 2$	$\sum e^2 / N-K$	

Total Variation	$\sum y^2 = 1480$	$n-1=5-1$ $=4$	$=\frac{97.5596}{2}$ $=48.7798$	

وبمقارنة F المحتسبة مع F الجدولية فإن النموذج المقدر معنوي.

6. اختبار معنوية المعلمات المقدرة:

يستخدم اختبار (t–Statistic) لمعرفة معنوية المعلمات المقدرة سواءٌ في نموذج الانحدار البسيط أم المتعدِّد إذا كان عدد المشاهدات يقل عن (30) مشاهدة. واختبار (Z) إذا كانت أكثر من ذلك. ويستخدم اختبار (t) وفق الآتي بإعادة كتابة المعادلة (4.20).

$$var - con\hat{\beta}_i = \sigma_u^2 (X'X)^{-1} \ldots \ldots (4.46)$$

فإذا كانت البيانات المستخدمة هي البيانات الحقيقية فتمثل عناصر القطر الرئيسي تباين المعلمات المقدرة بما فيها الحد الثابت وعلى التوالي والعناصر خارج القطر الرئيسي تمثل التباين المشترك. وبأخذ الجذر التربيعي لها نحصل على الخطأ المعياري للمعلمات ومنها نحسب اختبار (t) وكالاتي:

$$t_{\hat{\beta}_i} = \frac{\hat{\beta}_i - \beta_i}{S.e_{\hat{\beta}_i}} \ldots \ldots \ldots (4.47)$$

وتكوين قاعدة القرار يتحدد وفق ما مر ذكره في الانحدار البسيط.

أما إذا كانت البيانات المستخدمة هي بالانحرافات فإن أخذ الجذر التربيعي لعناصر المصفوفة في المعادلة (4.36) تعطي الخطأ المعياري للمعلمات عدا الحد الثابت، وعلى معنويتها يحسب اختبار (t) ابتداء من المعلمة $(\hat{\beta}_i)$. أما الاختبار للحد الثابت $(\hat{\beta}_0)$ فالخطأ المعياري له يحسب بأخذ الجذر التربيعي للمعادلة (4.37).

$$Se_{\hat{\beta}_0} = \sqrt{\sigma_u^2 \left[\bar{X}_i'(x'x)^{-1}\bar{X}_i + \frac{1}{n}\right]} \ldots \ldots \ldots \ldots (4.48)$$

$$t_{\hat{\beta}_i} = \frac{\hat{\beta}_0 - \beta_0}{S.e_{\hat{\beta}_0}} \ldots \ldots \ldots (4.49)$$

وبالرجوع إلى المثال السابق وإعادة كتابة $(x'x)^{-1}$:

$$(x'x) = \begin{bmatrix} \frac{29.7}{1044} & \frac{21}{1044} \\ \frac{21}{1044} & \frac{50}{1044} \end{bmatrix}$$

$$\sigma_u^2 = \frac{\sum e^2}{N-K} = 48.7798$$

$$\sigma_u^2 x'x = 48.7798 \begin{bmatrix} \frac{29.7}{1044} & \frac{21}{1044} \\ & \frac{50}{1044} \end{bmatrix} = \begin{bmatrix} 1.3877 & 0.9812 \\ & 2.3362 \end{bmatrix}$$

$$t_{\hat{\beta}_1} = \frac{5.707}{\sqrt{1.3877}} = \frac{5.707}{1.178} = 4.845$$

$$t_{\hat{\beta}_2} = \frac{1.2069}{\sqrt{2.3362}} = \frac{1.2069}{1.5285} = 0.7896$$

$$Se_{\hat{\beta}_0} = \sqrt{48.7798\left[(8 \quad 8.4)\begin{bmatrix} 50 & -21 \\ -21 & 29.7 \end{bmatrix}\begin{bmatrix} 8 \\ 8.4 \end{bmatrix} + \frac{1}{5}\right]}$$

$$Se_{\hat{\beta}_0} = \sqrt{48.7798\left[[223.6 \quad 81.48]\begin{bmatrix} 8 \\ 8.4 \end{bmatrix} + \frac{1}{5}\right]}$$

$$Se_{\hat{\beta}_0} = \sqrt{120643.7623 + \frac{1}{5}} = \sqrt{120643.9623}$$

$$Se_{\hat{\beta}_0} = 347.338$$

$$t_{\hat{\beta}_0} = \frac{6.20604}{347.338} = 0.0179$$

وبمقارنة القيم المحتسبة مع القيمة الجدولية البالغة 2.35 بمستوى معنوية 5% ودرجات حرية 3 نجد أن المعلمتين $\hat{\beta}_2 , \hat{\beta}_1$ غير معنويتين.

7. التباين والخطأ المعياري لمقدرات OLS:

مصفوفة التباين والتباين المشترك لمقدرات نموذج الانحدار الخطي المتعدِّد:

في حالة K من المتغيرات المستقلة فإن التباين إلى B_k يمكن الحصول عليه كما يلي:

$$\hat{B} = (X'X)^{-1}X'Y$$

$$\because Y = XB + U$$

وبالتعويض في المعادلة أعلاه نحصل:

$$\hat{B} = (X'X)^{-1}X'(XB + U)$$

ويمكن إعادة كتابة هذه المعادلة كما يلي:

$$\hat{B} = (X'X)^{-1}X'XB + (X'X)^{-1}X'U$$

ولما كانت: $(X'X)^{-1}X'X = 1$

فإن: $\hat{B} = B + (X'X)^{-1}X'U$

وبأخذ التوقع لطرفي المعادلة:

$$\mathrm{E}(\hat{B}) = B + E[(X'X)^{-1}X'U]$$
$$= B + (X'X)^{-1}X'E(U)$$
$$\because E(U) = 0$$
$$\because E(\hat{B}) = 0$$

إذًا $\hat{B}$ هي عبارة عن مقدار غير متحيز إلى B الحقيقية. ويمكن الحصول على تباين قيمة المعلمة المقدرة $\hat{B}$ حيث يؤدي ذلك إلى مصفوفة التباين والتباين المشترك الخاصة بالمتجه $\hat{B}$:

$$\hat{B} - E(\hat{B}) = \hat{B} - B = (X'X)^{-1}X'U$$
$$var(\hat{B}) = E\left\{[\hat{B} - E(\hat{B})][\hat{B} - E(\hat{B})]'\right\}$$
$$\because E(\hat{B}) = B$$
$$var\,(\hat{B}) = E\left\{(\hat{B} - B)(\hat{B} - B)'\right\} = E\{[(X'X)^{-1}X'U][(X'X)^{-1}X'U]'\,\}$$
$$= E\{(X'X)^{-1}X'UU'(X'X)^{-1}\}$$
$$= (X'X)^{-1}X'E(U'U)XU'(X'X)^{-1}$$

ولما كانت:
$$E(UU') = \sigma^2 In$$
$$var(\hat{B}) = (X'X)^{-1}X'.\sigma^2 In.X(X'X)^{-1}$$

وبإعادة الترتيب نحصل:
$$var(\hat{B}) = \sigma^2 In(X'X)^{-1}X'X(X'X)^{-1}$$

مصفوفة الوحدة:
$$\therefore\ (X'X)^{-1}X'X = 1$$
$$var(\hat{B}) = \sigma^2(X'X)^{-1}$$

هذا يعني أن قيمة تباين أي عنصر من عناصر المتجه $(\hat{B})$ هو عبارة عن حاصل ضرب قيمة σ^2 بما يقابلها من العناصر الواقعة على القطر الرئيسي لمقلوب المصفوفة $(X'X)^{-1}$. أما التباين المشترك بين أي اثنين من عناصر المتجه $(\hat{B})$ فهو عبارة عن حاصل ضرب σ^2 بالعنصر المقابل لها والواقع خارج القطر الرئيسي لمقلوب المصفوفة $(X'X)^{-1}$. ويمكن توضيح ذلك لنموذج الانحدار الخطي المتعدِّد الذي يحتوي على متغيرين مستقلين، وكما يلي:

$$var(\hat{B}) = \sigma^2(X'X)^{-1}$$
$$= \frac{\sigma^2}{n\sum X^2 - (\sum X)^2}\begin{bmatrix} \sum X^2 & \sum X \\ -\sum X & n \end{bmatrix}$$
$$= \begin{bmatrix} \dfrac{\sigma^2 \sum X^2}{n\sum X^2 - (\sum X)^2} & \dfrac{-\sigma^2 \sum X}{n\sum X^2 - (\sum X)^2} \\ \dfrac{-\sigma^2 \sum X}{n\sum X^2 - (\sum X)^2} & \dfrac{n\sigma^2}{n\sum X^2 - (\sum X)^2} \end{bmatrix}$$

$$var(\hat{B}_0) = \frac{\sigma^2 \sum X^2}{n\sum X^2 - (\sum X)^2} = \sigma^2 \left[\frac{1}{n} + \frac{\bar{X}^2}{\sum X^2}\right]$$

$$var(\hat{B}_1) = \frac{n\sigma^2}{n\sum X^2 - (\sum X)^2} = \frac{\sigma^2}{\sum X^2}$$

$$Cov(\hat{B}_0, \hat{B}_1) = \frac{-\sigma^2 \sum X}{n\sum X^2 - (\sum X)^2} = \frac{-\sigma^2 \bar{X}}{\sum X^2}$$

أما الصيغة التقديرية لتباين الخطأ العشوائي σ^2، (S^2e) فيمكن الوصول إليها كالآتي:

$$S^2e = \frac{e'e}{n-k-1} \ldots\ldots\ldots (10.4)$$

$$e'e = (Y - \hat{Y})'(Y - \hat{Y})$$

$$\because \hat{Y} = X\hat{B}$$

$$e'e = (Y - X\hat{B})'(Y - X\hat{B})$$

$$= Y'Y - Y'X\hat{B} - \hat{B}'X'Y + \hat{B}'X'X\hat{B} \ldots\ldots (11.4)$$

وبما أن الحدين الثاني والثالث يمثلان قيمة واحدة كما وأن كلاً منهما يمثل مبدلة للآخر فإن:

$$Y'X\hat{B} = (Y'X\hat{B})' = Y'X\hat{B}'$$

عليه يمكن إعادة كتابة المعادلة (11.4) بالشكل التالي:

$$e'e = Y'Y - 2\hat{B}'X'Y + \hat{B}'X'X\hat{B} \ldots\ldots (12.4)$$

$$\because \hat{B} = (X'X)^{-1}X'Y$$

$$\therefore (X'X)\hat{B} = X'Y$$

$$e'e = Y'Y - 2\hat{B}'X'Y + \hat{B}'X'Y \ldots\ldots\ldots (13.4)$$

$$e'e = Y'Y - \hat{B}'X'Y \ldots\ldots\ldots\ldots\ldots\ldots\ldots (14.4)$$

وبتعويض هذه القيمة في البسط من المعادلة (10.4) نحصل:

$$S^2e = \frac{Y'Y - \hat{B}'X'Y}{n-k-1}$$

8. اختبار فرضيات النموذج الخطي المتعدِّد:

يهدف هذا المبحث إلى توسيع معارفنا الأساسية لنموذج الانحدار وذلك بإجراء اختبار معنوية الانحدار المتعدِّد والمقدر باستخدام توزيع اختبار إحصاؤه F ومقارنته باختبار t ومن ثم تقييم كفاءة الأداء العام لنموذج الانحدار المتعدِّد R^2 ومقارنته بمعامل التحديد المقدر المعدل $\bar{R}^2$ ، وكذلك اختبار العلاقة بين F و R^2 من خلال جدول تحليل التباين، ANOVA ، ثم علاقة R^2 بقيمة المتغير العشوائي، $\sum e_i^2$. (عبد القادر عطية، 2004 صفحة 340)

9. اختبار معنوية المعالم (t):

يستخدم اختبار (t) لتقييم معنوية تأثير المتغيرات المستقلة $x_1, x_2, \dots x_k$ في المتغير التابع y في نموذج الانحدار المتعدِّد، وكما ذكرنا سابقاً عند تناول اختبار (t) في نموذج الانحدار الخطي البسيط، أنه يعتمد على نوعين من الفروض:

فرضية العدم: H_0 : $B_1 = B_2 = B_3 = \cdots .. = B_k = 0$

الفرضية البديلة: H_1 : $B_1 \neq B_2 \neq B_3 \neq \cdots .. \neq B_k = 0$

وبعد احتساب قيمة (t) تقارن مع قيمتها الجدولية لتحديد قبول أو رفض فرضية العدم ومن ثم تقييم معنوية معلمات النموذج المقدر والصيغة الرياضية لهذا الاختبار يمكن بيانها كما يلي:

أ – بالنسبة الى $\hat{B}_1$:

$$t_{\hat{B}_1} = \frac{\hat{B}_1}{S_{\hat{B}_1}}$$
$$S_{\hat{B}_1} = \sqrt{S^2_{\hat{B}_1}}$$
$$S^2_{\hat{B}_1} = var(\hat{B}_1) = S^2e\, a_{11}$$
$$var(\hat{B}_1) = S^2e(x'x)^{-1}$$
$$S^2e = \frac{e'e}{n-k-1} = \frac{Y'Y - \hat{B}'X'Y}{n-k-1} = \frac{\sum y^2 - \left(\hat{B}_1 \sum x_1 y + \hat{B}_2 \sum x_2 y\right)}{n-k-1}$$

ب– بالنسبة الى $\hat{B}_2$:

$$t_{\hat{B}_2} = \frac{\hat{B}_2}{S_{\hat{B}_2}}$$
$$S_{\hat{B}_2} = \sqrt{S^2_{\hat{B}_2}}$$
$$S^2_{\hat{B}_2} = var(\hat{B}_2) = S^2e\, a_{22}$$
$$S^2e = \frac{e'e}{n-k-1}$$

10. معامل التحديد المتعدِّد Multiple Coefficient of Determination (R):

يعد مؤشر أساس في تقييم مدى معنوية العلاقة بين المتغير التابع (Y)، والمتغيرات المستقلة (X_k)، إذ $(k = 1,2, \dots, k)$ بعبارة أخرى هو مقياس يوضح نسبة مساهمة المتغيرات المستقلة في تفسير التغير الحاصل في المتغير التابع. ويمكن اشتقاقه باستخدام المصفوفات بالانحرافات كالآتي:

$$\because y = x\hat{B} + e$$
$$e = y - x\hat{B}$$
$$e'e = \left(y - x\hat{B}\right)'\left(y - x\hat{B}\right)$$
$$e'e = y'y - y'x\hat{B} - x'\hat{B}'y + \hat{B}'x'x\,\hat{B}$$

وبما أن الحدين الثاني والثالث يمثلان قيمة واحدة كما وأن كلاً منهما يمثل مبدلاً للآخر فإن:

$$\therefore e'e = y'y - 2\hat{B}'x'y + \hat{B}'x'x\,\hat{B}$$
$$\because \hat{B}(x'x)^{-1}x'y$$
$$(x'x)\hat{B} = x'y$$
$$e'e = y'y - 2\hat{B}'x'y + \hat{B}'x'y$$
$$e'e = y'y - \hat{B}'x'y$$

بذلك يمكن كتابة معادلة الانحرافات الكلية كالآتي:

$$y'y = \hat{B}'x'y - e'e$$

إذ أن:

$y'y$: تمثل الانحرافات الكلية.

$\hat{B}'x'y$: تمثل الانحرافات الموضحة من قبل خط الانحدار.

$e'e$: تمثل الانحرافات غير الموضحة.

وبما أن معامل التحديد R^2 عبارة عن نسبة الانحرافات الموضحة من قبل خط الانحدار إلى الانحرافات الكلية Total variation ، فإنه يمثل نسبة مجموع مربعات التغير في المتغيرات المستقلة إلى مجموع المربعات الكلية:

$$\therefore R^2 = \frac{\hat{B}'x'y}{y'y} = \frac{\hat{B}'x'y}{\sum y^2}$$
$$or\ :\quad R^2 = 1 - \frac{e'e}{y'y - n\bar{Y}^2}$$
$$or:\quad R^2 = \frac{\hat{B}_1\sum x_1y + \hat{B}_2\sum x_2y}{\sum y^2}$$

إن إضافة متغيرات مستقلة جديدة إلى المعادلة يؤدي الى رفع قيمة R^2 ، وذلك لثبات قيمة المقام وتغير قيمة البسط بمقدار ($\hat{B}xy$) غير أن الاستمرار بإضافة المتغيرات المستقلة سيؤدي إلى انخفاض درجات الحرية $(n-k-1)$، مما يتطلب استخراج معامل التحديد المعدل أو المصحح R^2 وعلى النحو الآتي:

$$\bar{R}^2 = 1 - \left[(1-R^2)\frac{n-1}{n-k-1}\right]$$

11. اختبار إحصائية F-Statistics F:

.... هذا الاختيار معرفة مدى معنوية العلاقة الخطية بين المتغيرات المستقلة $X_1, X_2, \dots, X_i$ على المتغير التابع Y ، وكما هو الحال في الانحدار البسيط فإنه يعتمد على نوعين من الفروض:

فرضية العدم H_0: وتنص على انعدام العلاقة بين كل متغير من المتغيرات المستقلة $(X_1, X_2, \dots, X_i)$وبين المتغير التابع Y، أي:

$$H_0:\ \hat{B}_1 = \hat{B}_2 = \cdots . = \hat{B}_k = 0$$

الفرضية البديلة H_1: وتنص على وجود علاقة معنوية بين المتغيرات المستقلة والمتغير التابع، أي:

$$H_0:\ \hat{B}_1 \neq \hat{B}_2 \neq \cdots . \neq \hat{B}_k \neq 0$$

والصيغة الرياضية لهذا الاختبار هي:

$$F = \frac{\hat{B}'x'\,y/k}{e'e/n-k-1}$$

$$or:\quad F = \frac{R^2/k}{1-R^2/n-k-1}$$

وبعد احتساب قيمة F نقارن مع قيمتها بالجدولية بدرجة حرية (k) و $(n-k-1)$ للبسط والمقام ولمستوى معنوية معين. فإذا كانت القيمة المحتسبة أكبر من القيمة الجدولية نرفض H_0 ونقبل H_1، أي أن العلاقة المدروسة معنوية، وهناك على الأقل متغير مستقل واحد من المتغيرات X_k ذو تأثير في Y. أما إذا كنت القيمة المحتسبة أصغر من الجدولية فإن ذلك يعني قبول H_0 أي أن العلاقة الخطية المدروسة غير معنوية، أي أنه ليس ثمة تأثير من أي متغير من المتغيرات المستقلة على المتغير التابع.

12. تحليل جدول التباين، ANOVA

لغرض الوقوف على تأثير كل من X_1 ، X_2 في المتغير التابع Y، لابد من عمل جدول تحليل التباين لبيان أثر المتغيرين المستقلين X_1 ، X_2 في النموذج.

جدول تحليل التباين

مصدر التباين	مجموع مربعات الخطأ	درجات الحرية	متوسط مربعات الخطأ	اختبـار F
الانحراف الموضح من قبل X_1وX_2	$\hat{B}'x'y$ 1201.622859	K_2	$\hat{B}'x'\,y/k$ 600.8114295	$F_1 = \dfrac{\hat{B}'x''\,y/k}{e'\,e/n-k-1} = \dfrac{600.8114295}{12.06285679} = 49.8067$
الانحرافات غير الموضحة	$e'e$ 72.37714076	$n-k-1$ 6	$e/\ n-k-1$ 12.06285679	
الانحراف الكلي	$y'y$ 1274	$n-k$ 8		

ولمعرفة تأثير كل متغير مستقل في المتغير التابع بصورة منفردة فإننا نختبر خلال المرحلة الأولى تأثير المتغير X_2 بصورة مستقلة في Y. وفي المرحلة الثانية نختبر تأثير المتغير X_2 بصورة مستقلة في Y وبالرجوع إلى مثالنا نختبر ما يلي:

1.12 تأثير عنصر الدخل X_i في الاستيرادات Y:

لغرض اختبار التأثير المستقل لعنصر الدخل (X_i) في دالة الاستيرادات (Y) يجب معرفة مقدار الزيادة المتحققة في قيمة مجموع مربعات الانحرافات الموضحة من قبل خط انحدار المتغير التابع Y على المتغير المستقل X_2 نتيجة إضافة المتغير X_i الى الدالة، ويتم ذلك بافتراض نموذج يتضمن المتغير X_2 أي أن:

$$Y_i = B_0 + B_2X_2 + U_i$$

$$\hat{Y} = \hat{B}_0 + \hat{B}_2X_2$$

$$\hat{B}_2 = \frac{\sum x_2y}{\sum x_2^2} = \frac{-83}{648} = -0.128086419$$

أما مجموع مربعات الانحرافات الناجمة عن المتغير X_2 فهي:

$$\hat{B}_2\sum x_2y = (= -0.128086419)(-83) = 10.63117284$$

أما التأثير الذي يضيفه المتغير المستقل X_1 إلى الدالة فهو:

$$\hat{B}_1\sum x_1y = \hat{B}'X'y - \hat{B}_2\sum x_2y$$
$$= 1201.622859 - 10.63117284 = 1190.991686$$

ولاختبار مدى معنوية هذا التأثير المضاف من قبل المتغير المستقل X_1 ننظم جدول تحليل التباين الآتي:

تحليل التأثير المستقل للمتغير المستقل X_1 في النموذج

مصدر التباين	مجموع مربعات الخطأ	درجات الحرية	متوسط مربعات الخطأ	اختبار F
الانحراف الموضح من قبل X_2	10.63117284	1	1190.9911686	$F_1 = \frac{1190.9914686}{12.06285679} = 9873214171$
الانحراف الموضح من قبل X_1	1190.991686	1		
الانحراف الموضح من قبل X_1 و X_2	1201.622859	2		
الانحرافات غير الموضحة	72.37714076	6	12.06285679	
الانحرافات الكلية	1274	8		

وبمقارنة قيمة (F_1) المحتسبة والبالغة (98.73) مع مثيلتها الجدولية عند مستوى معنوية (5%) ودرجة حرية (2، 6) للبسط والمقام والبالغة (5.14) والتي يتضح أنها أكبر من الجدولية عليه ترفض فرضية العدم وتقبل الفرضية البديلة، مما يدل على وجود تأثير معنوي للمتغير المستقل X_1 الذي يمثل الدخل القومي على المتغير التابع Y الذي يمثل الاستيرادات.

2.12 تأثير عنصر السعر X_2 في الاستيرادات Y:

لبيان أثر المتغير المستقل X_2 في الدالة نفترض نموذجًا يتضمن المتغير المستقل X_1، أي:

$$Y_i = B_0 + B_1 X_1 + U_i$$
$$\hat{Y} = \hat{B}_0 + \hat{B}_1 X_1$$
$$\hat{B}_1 = \frac{\sum x_1 y}{\sum x_1^1} = \frac{881}{650} = 1.355384615$$

أما مجموع مربعات الانحرافات الناجمة عن المتغير X_1 فهي:

$$\hat{B}_1 \sum x_1 y = (1.355384612)(881) = 1194.093846$$

أما التأثير الذي يضيفه المتغير المستقل X_2 إلى الدالة فهو :

$$\hat{B}_2 \sum x_2 y = \hat{B}' x' y - \hat{B}_1 \sum x_1 y$$

$$= 1201.622859 - 1194.093846 = 7.529013$$

ولاختبار مدى معنوية هذا التأثير المضاف من قبل المتغير المستقل X_2 ننظم جدول تحليل التباين الآتي:

تحليل التأثير المستقل للمتغير المستقل X_2 في النموذج

مصدر التباين	مجموع مربعات الخطأ	درجات الحرية	متوسط مربعات الخطأ	اختبار F
الانحراف الموضح من قبل X_1	1194.093846	1	7.529013	$F_1 = \frac{7.529013}{12.06285679} = 0.624148419$
الانحراف الموضح من قبل X_2	7.529013	1		
الانحراف الموضح من قبل X_1 و X_2	1201.622859	2		
الانحرافات غير الموضحة	72.37714076	6	12.06285679	
الانحرافات الكلية	1274	8		

وبمقارنة قيمة (F_2) المحتسبة والبالغة (0.62) مع مثيلتها الجدولية عند مستوى معنوية (5%) ودرجة حرية (2، 6) للبسط والمقام والبالغة (5.14) والتي يتضح بأنها أقل من الجدولية عليه نقبل فرضية العدم، مما يدل بأن المتغير المستقل X_2 لا يمارس تأثيرًا معنويًا على المتغير التابع Y.

وعليه نستنتج بأن المتغير المستقل X_1 الذي يمثل الدخل القومي يمارس تأثيرًا معنويًا على المتغير التابع Y الذي يمثل الاستيرادات. في حين أن المتغير المستقل X_2 الذي يمثل السعر لا يمارس تأثيرًا معنويًا على Y ومن ثم يجب استبعاده من النموذج المدروس واعتماد النموذج الذي يحتوي على المتغير X_1 ذو التأثير المعنوي في المتغير Y والذي ندرجه أدناه:

$$Y_i = B_0 + B_1 X_1 + U_i$$

حيث يتم تقديره وتقييمه على غرار النموذج الوارد في الفصل الخاص بالانحدار الخطي البسيط وكما يلي:

$$\hat{B}_1 = \frac{\sum x_1 y}{\sum x_1^2} = \frac{881}{650} = 1.355384615$$

$$\hat{B}_0 = \bar{Y} - \hat{B}_1 \bar{X}$$
$$= 117 - (1.355384615)(113)$$
$$= 117 - 153.1584615$$
$$= -36.1584645$$
$$\hat{Y} = \hat{B}_0 + \hat{B}_1 X_i$$
$$\hat{Y} = -36.1584615 + 1.355384615\ X_i$$

وتعني بأن زيادة المتغير المستقل X والذي يمثل الدخل القومي بمقدار وحدة واحدة تسبب زيادة في المتغير التابع Y الذي يمثل الاستيرادات بمقدار (1.355) وحدة.

$$R^2 = \frac{\hat{B}_1 \sum x_1 y}{\sum y^2} = \frac{(1.355384615)(881)}{1274} = \frac{1194.093846}{1274} = 0.937279313$$
$$= 93.72\%$$

$$F = \frac{R^2/k}{1 - R^2/n-k-1} = \frac{0.93727931\,3/1}{1 - 0.93727931\,3/9-1-1} = \frac{0.937279313}{0.008960098}$$
$$= 104.6059221$$

ولاختبار معنوية المعلمات $\hat{B}_1$ و $\hat{B}_0$ نحتاج البيانات الآتية:

Y	$\hat{Y}$	e_i	e_i^2
100	99.38	0.62	0.3844
106	104.8015385	1.1984615	1.436309967
107	107.5123077	−0.5123077	0.262459179
120	114.2892308	5.7107692	32.61288486
110	114.2892308	−4.2892308	18.39750086
116	119.7107692	−3.7107692	13.7698086
124	126.4876923	−2.4876923	6.188612979
133	131.9092308	1.0907692	1.18977448
137	134.62	2.38	5.6644
$\sum Y = 1053$	$\sum \hat{Y} = 1053$	$\sum e_i = 0$	79.90615335

بالنسبة لـ $\hat{B}_1$:

$$S^2 e_i = \frac{\sum e_i^2}{n-2} = \frac{79.90615335}{9-2} = 11.41516476$$

$$S_{\widehat{B}_1} = \frac{S^2 e_i}{\sum x_i^2} = \frac{11.41516476}{650} = 0.017561791$$

$$S_{\widehat{B}_1}^2 = \sqrt{S_{\widehat{B}_1}^2} = \sqrt{0.017561791} = 0.132520911$$

$$t_{\widehat{B}_1} = \frac{\widehat{B}_1}{S_{\widehat{B}_1}} = \frac{1.355384615}{0.132520911} = 10.22770372$$

بالنسبة لـ $\widehat{B}_0$:

$$S^2 e = 11.41516476$$

$$S_{\widehat{B}_0}^2 = S^2 e\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right] = 11.41516476\ \left[\frac{1}{9} + \frac{(113)^2}{650}\right]$$

$$= 11.41516476[0.111111111 + 19.64461538] = 225.5148729$$

$$S_{\widehat{B}_0} = \sqrt{S_{\widehat{B}_0}^2} = \sqrt{225.5148729} = 15.01715262$$

$$t_{\widehat{B}_0} = \frac{\hat{B}_0}{S_{\widehat{B}_0}} = \frac{-36.15846415}{15.01715262} = -2.407810749$$

وعليه فإن الصيغة التقديرية للنموذج المدروس تكون كما يلي:

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

$$\hat{Y} = -36.1584615 + 1.355384615 X_i$$

$$(2.407) \qquad (10.227)$$

$$R^2 = \%93.72 \quad F = 104.60$$

13.قياس حدود الثقة:

لاحتساب حدود الثقة لأية مشاهدة (نقطة) من مشاهدات خط الانحدار للمجتمع أو بعبارة أخرى لحساب القيمة المتوسطة الحقيقية إلى Y عند مستوى معنوية معين للمتغير المستقل في النموذج. نفترض بأن النقطة المراد تقدير حدود ثقتها هي $F(Y_0)$. ولتقدير المجال الذي يمكن أن تقع فيه قيمة $F(Y_0)$ المقابلة لتشكيلة معينة من قيم المتغيرات المستقلة (k) يجب اشتقاق متباينة القيمة $F(Y_0)$.

$$Y_0 = [1 \quad X_{01} \quad X_{02} \quad \dots X_{0k}]$$

$$\hat{Y}_0 = [1 \quad X_{01} \quad X_{02} \quad \dots X_{0k}] \begin{bmatrix} \hat{B}_0 \\ \hat{B}_1 \\ . \\ . \\ \hat{B}_k \end{bmatrix}$$

$$\hat{Y}_0 = \hat{B}_0 + \hat{B}_1 X_{01} + \dots + \hat{B}_k X_{0k}$$

وباختصار:

$$\hat{Y}_0 = X_0 \hat{B}$$

ولغرض اشتقاق المتباينة الخاصة بتقدير فترات حدود الثقة للقيمة $E(Yo)$ يجب اشتقاق وسط وتباين القيمة $(\hat{Y}_0)$ وكالآتي:

لإيجاد الوسط فإننا نأخذ القيمة المتوقعة لـ $(\hat{Y}_0)$:

$$E(\hat{Y}_0) = E(X_0\hat{B})$$
$$E(\hat{Y}_0) = X_0E(\hat{B})$$
$$\because E(\hat{B}) = B$$
$$\because E(\hat{Y}_0) = X_0B$$

ولإيجاد التباين:

$$var(\hat{Y}_0) = E\{\hat{Y}_0 - E(\hat{Y}_0)(\hat{Y}_0 - E(\hat{Y}_0)'\}$$
$$= E\left\{(\hat{Y}_0 - X_0B)(\hat{Y}_0 - X_0B)'\right\}$$
$$\therefore \hat{Y}_0 = X_0B$$
$$\therefore var(\hat{Y}_0) = E\left\{(X_0\hat{B} - X_0B)(\hat{Y}_0 - X_0B)'\right\}$$
$$\therefore var(\hat{Y}_0) = E\{(X_0\hat{B} - X_0B)(X_0B' - X_0B)'\}$$
$$= X_0\left\{E(\hat{B} - B)(\hat{B} - B)'\right\}X_0'$$
$$= \sigma^2X_0(X'X)^{-1}X_0'$$

وإذا رمزنا للقيمة التقديرية لتباين قيمة حدود الثقة $var(\hat{Y}_0)$ بالرمز $S^2(\hat{Y}_0)$ فإن:

$$S^2(\hat{Y}_0) = S^2X_0(X'X)^{-1}X_0'$$

وعليه فإن حدود الثقة للقيمة $E(Y_0)$ تكون:

$$E(\hat{Y}_0) = \hat{Y}_0 \pm t_{1/2}.S(\hat{Y}_0)$$
$$E(\hat{Y}_0) = X_0\hat{B} \mp t_{1/2}.S(\hat{Y}_0)$$

المحور الرابع:

الارتباط الجزئي، الازدواج الخطي وطرق اختيار المتغيرات التفسيرية

تمهيد:

يقيس معامل الارتباط الجزئي صافي الارتباط بين المتغير التابع ومتغير مستقل بعد حذف التأثير المشترك أي مع تثبيت المتغيرات المستقلة الأخرى في النموذج. فمثلاً $\mathrm{r}_{YX_1.X_2}$ هو الارتباط الجزئي بين Y و X_1 بعد حذف تأثير X_2 من كل من Y و X_1:

المسألة 7−14 (أ):

$$\mathrm{r}_{YX_1.X_2}=\frac{\mathrm{r}_{YX_1}-\mathrm{r}_{YX_2}r_{X_1X_2}}{\sqrt{1-r_{X_1X_2}^2}\sqrt{1-r_{YX_2}^2}} \qquad (14-7)$$

$$\mathrm{r}_{YX_2.X_1}=\frac{\mathrm{r}_{YX_2}-\mathrm{r}_{YX_1}r_{X_1X_2}}{\sqrt{1-r_{X_1X_2}^2}\sqrt{1-r_{YX_2}^2}} \qquad (15-7)$$

حيث r_{YX_1} معامل الارتباط البسيط بين Y و X_1، ويعرف r_{YX_2} و $r_{X_1X_2}$ على نفس الخط. وتتراوح معاملات الارتباط الجزئية بين 1−،1+ (كما هو الحال بالنسبة لمعاملات الارتباط البسيط)، ويكون لها نفس إشارة معلمة المجتمع المناظرة، وتستخدم لتحديد الأهمية النسبية للمتغيرات المفسرة المختلفة في الانحدار المتعدِّد.

مثال5: بالتعويض بقيم جدول 7−1، 7−2 في معادلة (6−18) لمعامل الارتباط البسيط، نحصل على:

$$r_{YX_1}=\frac{\sum x_1y}{\sqrt{\sum x_1^2}\sqrt{\sum y^2}}=\frac{956}{\sqrt{576}\sqrt{1,634}}\approx 0.9854$$

$$r_{YX_2}=\frac{\sum x_2y}{\sqrt{\sum x_2^2}\sqrt{\sum y^2}}=\frac{900}{\sqrt{504}\sqrt{1,634}}\approx 0.9917$$

$$r_{X_1X_2}=\frac{\sum x_2x_1}{\sqrt{\sum x_2^2}\sqrt{\sum x_1^2}}=\frac{524}{\sqrt{504}\sqrt{576}}\approx 0.9725$$

وعليه:

$$r_{YX_1..X_2}=\frac{r_{YX_1}-r_{YX_2}r_{X_1X_2}}{\sqrt{1-r_{X_1X_2}^2}\sqrt{1-r_{YX_2}^2}}=\frac{0.9854-(0.9917)(0.9725)}{\sqrt{1-0.9725^2}\sqrt{1-0.9917^2}}\cong 0.7023 or 70.23\%$$

و:

$$r_{YX_2.X_1}=\frac{r_{YX_2}-r_{YX_1}r_{X_1X_2}}{\sqrt{1-r_{X_1X_2}^2}\sqrt{1-r_{YX_1}^2}}=\frac{0.9917-(0.9854)(0.9725)}{\sqrt{1-0.9725^2}\sqrt{1-0.9854^2}}\cong$$

$$\cong 0.8434 \ or\ 84.34\%$$

وعليه، فإن X_2 أكثر أهمية من X_1 في تفسير التغير في Y.

مثال 6: يمكن تلخيص النتائج الكلية لمثال الحنطة – السماد – المبيد كالآتي:

$$\hat{Y} = 31.98 + 0.65X_1 + 1.10X_2$$

t ـــــ قيم (2.70) (4.11)

$$R^2 = 0.992 \qquad \bar{R}^2 = 0.989 \qquad F_{2,7} = 413.17$$

$$r_{YX_1.X_2} = 0.70 \qquad r_{YX_2X_1} = 0.84$$

وبالرغم من الحصول على النتائج عادة باستخدام الكمبيوتر، إلا أنه من المهم القيام بحل المسألة (يدويًا)، كما فعلنا لكي نفهم خطوات الحل بوضوح. وتعرض المسألة 7−22 عينة برنامج – كمبيوتر كامل يشرح بالكامل كيفية استخدام Statistical Package for the Social Sciences) SPSS وهو أكثر برامج الكمبيوتر شيوعًا في الاستخدام)، لانحدار متعدِّد في ثلاث متغيرات (دومينيك، 1983 صفحة 169).

مسائل محلولة:

1.النموذج الخطي ذو المتغيرات الثلاث:

7−1: (أ) اكتب معادلة نموذج الانحدار الخطي المتعدِّد لحالة متغيرين مستقلين أو مفسرين وحالة k متغير مستقل أو مفسر. (ب) اذكر فروض النموذج الخطي للانحدار المتعدِّد.

(أ) في حالة متغيرين مستقلين أو مفسرين، المعادلة هي:

$$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + u_i \qquad (1-7)$$

وفي حالة k متغير مستقل أو مفسر، المعادلة هي:

$$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + \cdots + b_kX_{ki} + u_i$$

حيث تمثل X_{2i}، على سبيل المثال المشاهدة التي ترتيبها i للمتغير المستقل X_2

(ب) الفروض الخمسة الأول لنموذج الانحدار الخطي المتعدِّد هي نفس فروض نموذج الانحدار البسيط OLS (انظر المسألة 6−1). أي أن الفروض الثلاثة الأول يمكن تلخيصها على النحو: $u_i \sim N(O,\sigma)$. الفرض الرابع هو $E_{(u_iu_j)} = 0$ عند $i \neq j$ ؛ والفرض الخامس هو $E(X_iu_j) = 0$. الفرض الإضافي الوحيد المطلوب لنموذج الانحدار الخطي المتعدِّد OLS هو أنه لا توجد علاقة خطية تامة بين المتغيرات المستقلة $X_1, X_2, \ldots$ لأنه لو كان بين اثنين أو أكثر من المتغيرات المستقلة ارتباط خطي تام، لاستحالة حساب تقديرات معالم OLS لأن مجموعة المعادلات الطبيعية سوف تشتمل على معادلتين أو أكثر ليست مستقلة. أما إذا كان هناك ارتباط خطي كبير وليس تامًا بين التين أو أكثر من المتغيرات المفسرة، فإنه يمكن تقدير معالم OLS، ولكن لا يمكن عزل تأثير كل من المتغيرات المستقلة ذات الارتباط الخطي الكبير فيما بينها (أنظر قسم 9−1)

7-2: باستخدام طريقة OLS في حالة متغيرين مستقلين أو مفسرين، اشتق (أ) المعادلة الطبيعية (7-2)، (ب) المعادلة الطبيعية (7-3)، (ج) المعادلة الطبيعية (7-4). (القارئ غير الملم بالتفاضل يمكنه أن يتخطي هذه المسألة).

(أ) تشتق المعادلة الطبيعية (7-2) بإيجاد النهاية الصغرى للمقدار $\sum e_i^2$ بالنسبة إلى $\hat{b}_0$:

$$\frac{\partial e_i^2}{\partial \hat{b}_0} = \frac{\partial \sum\left(Y_i - \hat{b}_0 - \hat{b}_1 X_{1i} - \hat{b}_2 X_{2i}\right)^2}{\partial \hat{b}_0} = 0$$

$$-2\sum\left(Y_i - \hat{b}_0 - \hat{b}_1 X_{1i} - \hat{b}_2 X_{2i}\right) = 0 \qquad (2-7)$$

$$\sum Y_i = n\hat{b}_0 \pm \hat{b}_1 \sum X_{1i} \pm \hat{b}_2 \sum X_{2i}$$

(ب) وتشتق المعادلة (7-3) بإيجاد النهاية الصغرى للمقدار $\sum e_i^2$ بالنسبة إلى $\hat{b}_1$:

$$\frac{\partial \sum e_i^2}{\partial \hat{b}_1} = \frac{\partial \sum\left(Y_i - \hat{b}_0 - \hat{b}_1 X_{1i} - \hat{b}_2 X_{2i}\right)^2}{\partial \hat{b}_1} = 0$$

$$-2X_{1i}\sum\left(Y_i - \hat{b}_0 - \hat{b}_1 X_{1i} - \hat{b}_2 X_{2i}\right) = 0 \qquad (3-7)$$

$$\sum X_{1i} Y_i = \hat{b}_0 \sum X_{1i} + \hat{b}_1 \sum X_{1i}^2 + b_2 \sum X_{1i} X_{2i}$$

(ج) وتشتق المعادلة الطبيعية (7-4) بإيجاد النهاية الصغرى للمقدار $\sum e_i^2$ بالنسبة إلى $\hat{b}_2$:

$$\frac{\partial \sum e_i^2}{\partial \hat{b}_2} = \frac{\partial \sum\left(Y_i - \hat{b}_0 - \hat{b}_1 X_{1i} - \hat{b}_2 X_{2i}\right)^2}{\partial \hat{b}_2} = 0$$

$$-2X_{2i}\sum\left(Y_i - \hat{b}_0 - \hat{b}_1 X_{1i} - \hat{b}_2 X_{2i}\right) = 0 \qquad (4-7)$$

$$\sum X_{2i} Y_i = \hat{b}_0 \sum X_{2i} + \hat{b}_1 \sum X_{1i} X_{2i} + b_2 \sum X_{2i}^2$$

7-3: بالنسبة لنموذج الانحدار الخطى المتعدِّد ذي المتغيربن المستقلين، (أ) اشتق المعادلات الطبيعية باستخدام الانحرافات (إرشاد: ابدأ باشتقاق تعبير $\hat{y}_i$: يمكن للقارئ غير الملم بالتفاضل أن يتخطى هذا الجزء من المسألة). (ب) كيف يمكن اشتقاق المعادلات 7-5، 7-6، 7-7، لإيجاد $\hat{b}_1$، $\hat{b}_2$، $\hat{b}_0$:

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

(أ)

$$\bar{Y} = \hat{b}_0 + \hat{b}_1 \bar{X}_1 + \hat{b}_2 \bar{X}_2$$

بالطرح، نحصل على:

$$\hat{y}_i = \hat{Y}_i - \bar{Y} = \hat{b}_1 x_{i1} + \hat{b}_2 x_{2i}$$

وعليه:

$$e_i = y_i - \hat{y}_i = \hat{b}_1 x_{1i} - \hat{b}_2 x_{2i}$$

$$\sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum\left(y_i - \hat{b}_1 x_{1i} - \hat{b}_2 x_{2i}\right)^2$$

$$\frac{\partial \sum e_i^2}{\partial \hat{b}_1} = \frac{\partial \sum \left(y_i - \hat{b}_1 x_{1i} - \hat{b}_2 x_{2i}\right)^2}{\partial \hat{b}_1} = 0$$

$$-2x_{1i} \sum \left(y_i - \hat{b}_1 x_{1i} - \hat{b}_2 x_{2i}\right) = 0 \qquad (16-7)$$

$$\sum x_{1i} y_i = \hat{b}_1 \sum x_{1i}^2 + \hat{b}_2 \sum x_{1i} x_{2i} \qquad (16-7)$$

$$\frac{\partial \sum e_i^2}{\partial \hat{b}_2} = \frac{\partial \sum \left(y_i - \hat{b}_1 x_{1i} - \hat{b}_2 x_{2i}\right)^2}{\partial \hat{b}_2} = 0$$

$$-2x_{2i} \sum \left(y_i - \hat{b}_1 x_{1i} - \hat{b}_2 x_{2i}\right) = 0$$

$$\sum x_{2i} y_i = \hat{b}_1 \sum x_{1i} x_{2i} + \hat{b}_2 \sum x_{2i}^2 \qquad (17-7)$$

(ب) المعادلات (7–5)، (7–6) لحساب $\hat{b}_1$ و$\hat{b}_2$ على الترتيب، يتم الحصول عليها بحل معادلات (7–16)، (7–17) آنيًا، ويمكن دائمًا حساب $\hat{b}_1$ و$\hat{b}_2$ ، إلاَّ إذا كانت هناك علاقة خطية تامة بين X_1 وX_2 أو كان عدد المشاهدات عن كل متغير في النموذج 3 أو أقل. ويمكن حساب المعلمة $\hat{b}_0$ بالتعويض في معادلة (7–7) بقيم $\hat{b}_1$ و$\hat{b}_2$ (المحسوبة باستخدام معادلات (7–5)، (7–6))، وقيم $\bar{Y}$ و $\bar{X}_1$ و$\bar{X}_2$ (المحسوبة من معطيات المسألة).

7–4: بالنسبة لتحليل الانحدار المتعدِّد ذي المتغيرين المستقلين بين معنى (أ) b_0 (ب) b_1، (ج) b_2 (د) هل $\hat{b}_0$ و$\hat{b}_1$ و $\hat{b}_2$ BLUE؟

(أ) المعلمة b_0 هي الحد الثابت أو مقطع الانحدار وتعطى قيمة المتغير Y_i، عندما $X_{i1} = X_{2i} = 0$

(ب) تقيس المعلمة b_1 التغير في Y لكل وحدة تغير في X_1 مع إبقاء X_2 ثابتة. ومعلمة الميل b_1 هي معامل انحدار جزئي لأنها تناظر المشتقة الجزئية للمتغير Y بالنسبة إلى X_1، أي $\partial Y/\partial X_1$.

(ج) تقيس المعلمة b_2 التغير في Y لكل وحدة تغير في X_2 مع إبقاء X_1 ثابتة. ومعلمة الميل b_2 هي المعامل الجزئي الثاني للانحدار لأنها تناظر المشتقة الجزئية للمتغير Y بالنسبة إلى X_2، أي: $\partial Y/\partial X_2$

(د) حيث أنه يتم الحصول على $\hat{b}_0$ و $\hat{b}_1$ و $\hat{b}_2$ بطريقة OLS، فإنها أيضًا أفضل مقدرات خطية غير متحيزة (BLUE، انظر قسم 6–5). أي أن $E(\hat{b}_0) = b_0$، $E(\hat{b}_1) = b_1$ ، $E(\hat{b}_2) = b_2$، و $\hat{S}_{b_0}$، $\hat{S}_{b_1}$، $\hat{S}_{b_2}$ أصغر منها لأي مقدرات خطية غير متحيزة أخرى، ولما كان إثبات هذه الخصائص يمثل عبئًا ثقيلاً بدون استخدام جبر المصفوفات لذا لا نتناولها هنا.

7–5: جدول 7–3 هو امتداد للجدول 6–11 ويعطي دخل الفرد الحقيقي بآلاف الدولارات Y، مع نسبة القوة العاملة في الزراعة، X_1 ومتوسط سنوات التعليم للسكان فوق سن 25 سنة، X_2 لعدد 15 دولة متقدمة في

1981. (أ) أوجد معادلة انحدار المربعات الصغرى للمتغير Y على X_1 و X_2 (ب) فسر النتائج في (أ) وقارنها بنتائج المسألة 6−30.

جدول 3−7 دخل الفرد، القوة العاملة في الزراعة، وسنوات التعليم

n	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5
Y	6	8	8	7	7	12	9	8	9	10	10	1	9	10	1
X_1	9	10	8	7	10	4	5	5	6	8	7	4	9	5	8
X_2	8	13	11	10	12	16	10	10	12	14	12	16	14	0	2

جدول 7−4 مسودة لتقدير المعالم لبيانات جدول 7−3

n	y	X_1	X_2	y	x_1	x_2	x_1y	x_2y	x_1x_2	x_1^2	x_2^2
1	6	9	8	3-	2	4-	6-	12	8-	4	16
2	8	10	13	-1	3	1	-3	-1	3	9	1
3	8	8	11	-1	1	-1	-1	1	-1	1	1
4	7	1	10	-2	0	-2	0	4	0	0	4
5	7	10	12	-2	3	0	-6	0	0	9	0
6	12	4	16	3	-3	4	-9	12	-12	9	16
7	9	5	10	0	-2	-2	0	0	4	4	4
8	8	5	10	-1	-2	-2	2	2	4	4	4
9	9	6	12	0	-1	0	0	0	0	1	0
10	10	8	14	1	1	2	1	2	2	1	4
11	10	7	12	1	0	0	0	0	0	0	0
12	11	4	16	2	-3	4	-6	8	-12	9	16
13	9	9	14	0	2	2	-0	0	4	4	4
14	10	5	10	1	-2	-2	-2	-2	4	4	4
15	11	8	12	2	1	0	2	0	0	1	0
$n = 15$	$\sum Y = 139$ $\bar{Y} = 9$	$\sum X_1 = 105$ $\bar{X}_1 = 7$	$\sum X_2 = 180$ $\bar{X}_2 = 12$	$\sum y = 0$	$\sum x_1 = 0$	$\sum x_2 = 0$	$\sum x_1y = -28$	$\sum x_2y = 38$	$\sum x_1x_2 = -12$	$\sum x_1^2 = 60$	$\sum x_2^2 = 74$

(أ) يبين جدول 7−4 الحسابات اللازمة لتقدير معالم معادلة انحدار OLS للمتغير Y على المتغيرين X_1 و X_2:

$$\hat{b}_1 = \frac{(\sum x_1 y)(\sum x_2^2) - (\sum x_2 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2} = \frac{(-28)(74) - (38)(-12)}{(60)(74) - (-12)^2}$$
$$= \frac{-2{,}075 + 456}{4{,}440 - 144} \cong -0.38$$
$$\hat{b}_2 = \frac{(\sum x_2 y)(\sum x_1^2) - (\sum x_1 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2} = \frac{(38)(60) - (-28)(-12)}{(60)(74) - (-12)^2} = \frac{2{,}280 - 336}{4{,}440 - 144}$$
$$\cong 0.45$$
$$\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X}_1 - \hat{b}_2 \bar{X}_2 = 9 - (-0.38)(7) - (0.45)(12) = 9 + 2.66 - 5.40 \cong 6.26$$

وعليه فمعادلة OLS لانحدار Y على X_1 و X_2 هي:

$$\hat{Y}_i = 6.26 - 0.38X_{1i} + 0.45X_{2i}$$

(ب) تشير معادلة انحدار OLS المقدرة على أن مستوى دخل الفرد الحقيقي Y ، يرتبط عكسيًا مع نسبة القوة العاملة في الزراعة X_1، وطرديًا مع عدد سنوات التعليم للسكان فوق من 25 (كما قد يكون متوقعًا). بالتحديد تشير $\hat{b}_1$، إلى أن نقص نسبة القوة العاملة في الزراعة بمقدار 1% من إجمالي القوة العاملة سوف يصاحبه زيادة قدرها 380 دولارًا أمريكيًا في دخل الفرد مع تثبيت X_2 ولكن زيادة سنة واحدة في سنوات التعليم للسكان فوق سن 25 سنة يصاحبها زيادة في دخل الفرد قدرها 450 دولارًا أمريكيًا، مع تثبيت X_1. وعند: $X_{1i} = X_{2i}| = 0,$ $\hat{Y}_i = \hat{b}_0 = 6.26$. وبقدر ما اتضح أن X_2 معنوية إحصائيًا (انظر المسألة 8-7 (ب))، وبالتالي يجب أن تدخل في علاقة الانحدار، فقد اتضح أيضًا أن $\hat{b}_1 \cong -0.47$ السابق إيجادها في تمرين6-30، لا تكون تقديرًا موثوقًا للمعلمة b_1.

2. اختبارات معنوية تقديرات المعالم:

6-7: عرف (أ) σ_u^2 و s^2، (ب) تباين $\hat{b}_1$ وتباين $\hat{b}_2$، (ج) و، $s_{\hat{b}_1}^2$ و $s_{\hat{b}_2}^2$ ، (د) $s_{\hat{b}_1}$ و $s_{\hat{b}_2}$. (هـ) لماذا لا تكون b_0 عادة موضع اهتمام أساسي ؟

(أ)σ_u^2 هو تباين حدا لخطأ في العلاقة الحقيقية بين X_{1i} وX_{2i} وY_i ولكن $s^2 = \hat{\sigma}_u^2 = \sum e_i^2/(n-k)$

هي تباين البواقي وهي تقدير غير متحيز للتباين غير المعلوم σ_u^2. k هي عدد المعالم المقدرة. في حالة الانحدار المتعدِّد ذي المتغيرين، $k = 3$. وعليه $n - k = n - 3 = df$

(ب):
$$Var\hat{b}_1 = \sigma_u^2 \frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$$

بينما:
$$Var\hat{b}_2 = \sigma_u^2 \frac{\sum x_1^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$$

إن تبايني $\hat{b}_1$ و $\hat{b}_2$ (أو تقديراتها) مطلوبة لاختبار الفروض وتكوين فترات الثقة لكل من b_1 و b_2 .

(ج):

$$s_{\hat{b}_1}^2 = s^2 \frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2} = \frac{\sum e_i^2}{n-k} \frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$$

$$s_{\hat{b}_2}^2 = s^2 \frac{\sum x_1^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2} = \frac{\sum e_i^2}{n-k} \frac{\sum x_1^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$$

$s_{\hat{b}_1}^2$ و $s_{\hat{b}_2}^2$ هما على الترتيب، تقديران غير متحيزين لتباين b_1 وتباين b_2 غير المعلومين حيث أن σ_u^2 غير معلومة.

(د): $S_{\hat{b}_1} = \sqrt{s_{\hat{b}_1}^2}$ و $S_{\hat{b}_2} = \sqrt{s_{\hat{b}_2}^2}$ ، $S_{\hat{b}_1}$ و $S_{\hat{b}_2}$ هما على الترتيب، الانحراف المعياري لكل من $\hat{b}_2$ و $\hat{b}_2$ ويسميان بالأخطاء المعيارية.

(هـ) ما لم تتوفر مشاهدات كافية بالقرب من $X_{1i} = X_{2i} = 0$ فإن معلمة المقطع b_0 لا تكون عادة ذات أهمية أساسية ويمكن حذف اختبار معنوية الإحصائية الخاص بها ومعادلة (7−18) لتباين $\hat{\mathrm{b}}_0$ معادلة معقدة في الحساب ولهذا السبب أيضًا فمن النادر أن تذكر أو تستخدم:

$$Var\hat{\mathrm{b}}_0 = \sigma_u^2 = \frac{\sum X_1^2 \sum X_2^2 - (\sum X_1 X_2)^2}{n[\sum X_1^2 X_2^2 - (\sum X_1 X_2)^2] - \sum X_1 (\sum X_1 \sum X_2^2 - \sum X_2 \sum X_1 X_2) + \sum X_1 (\sum X_1 X_2 - \sum X_2 \sum X_1^2)} \quad \dots\dots\dots\dots (18-7)$$

ومع ذلك، ترد $S_{\hat{b}_0}$ أحيانًا في نتائج الكمبيوتر، ويمكن إجراء الاختبارات الإحصائية لمعنوية b_0 بسهولة.

<u>7−7 :</u>من بيانات جدول 7−3، أوجد (أ) S^2، (ب) $s_{\hat{b}_1}^2$ و $S_{\hat{b}_1}$ (ج) $s_{\hat{b}_2}^2$ و $S_{\hat{b}_2}$

(أ) الحسابات اللازمة لإيجاد S^2 موضحة في جدول 7−5، وهو امتداد لجدول 7−4. وقد تم الحصول على قيم $\hat{Y}_i$ بالتعويض بقيم X_{i1} و X_{2i} في معادلة انحدار OLS المقدرة السابق إيجادها في المسألة 7−5(أ).

$$S^2 = \hat{\sigma}_u^2 = \frac{\sum e_i^2}{n-k} = \frac{12.2730}{15-3} \cong 1.02$$

جدول 7−5 انحدار دخل الفرد: حسابات اختبار معنوية المعالم

الدولة	Y	X_1	X_2	$\hat{Y}$	e	e^2
1	6	9	8	6.44	−0.44	0.1936
2	8	10	13	8.31	−0.31	0.0961
3	8	8	11	8.17	−0.17	0.0289
4	7	7	10	8.10	−1.10	1.2100
5	7	10	12	7.86	−0.86	0.7396
5	7	10	12	7.86	−0.86	0.0036
7	9	5	10	8.86	0.14	0.0196

8	8	5	10	8.86	−0.86	0.7396
9	9	6	12	9.38	−0.38	0.1444
10	10	8	14	9.52	0.48	0.2304
11	10	7	12	9.00	1.00	1.0000
12	11	4	16	11.94	−0.94	0.8836
13	9	9	14	9.14	−0.14	0.0196
14	10	5	10	8.86	1.14	1.2996
15	11	8	12	8.62	2.38	5.6644
n = 15					$\sum e = 0$	$\sum e^2 = 12.2730$

(ب) باستخدام قيمة S^2 السابق إيجادها في (أ) وقيم جدول 7−4، نحصل على:

$$s_{\hat{b}_1}^2 = S^2 \frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2} \cong 1.02 \frac{74}{(60)(74) - (-12)^2} \cong 0.02$$

$$s_{\hat{b}_1} \cong \sqrt{0.02} \cong 0.14$$

(ج)

$$s_{\hat{b}_1}^2 = s^2 \frac{\sum x_1^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2} \cong 1.02 \frac{60}{(60)(74) - (-12)^2} \simeq 0.01$$

$$s_{\hat{b}_2} \cong \sqrt{0.01} \cong 0.10$$

7−8: اختبر عند مستوى معنوية 5% كل من (أ) b_1 و (ب)b_2 في المسألة 7−5 (أ).

(أ)

$$t_1 = \frac{\hat{b}_1 - b_1}{s_{\hat{b}_1}} = \frac{-0.38-0}{0.14} \cong 2.71$$

وحيث أن قيمة t_1 المطلقة تتجاوز القيمة الجدولية $t = 2.179$ (من ملحق5) بمستوى معنوية 5% اختبار ذو ذيلين)، $n - k = 15 - 3 = 12df$ فإننا نستنتج أن b_1 معنوية إحصائيًا عند مستوى معنوية 5% (أي أننا لا نستطيع أن نرفض H_1، بأن $b_1 \neq 0$).

(ب)

$$t_2 = \frac{\hat{b}_2 - b_2}{S_{\hat{b}_2}} \cong \frac{0.45-0}{0.10} \cong 4.50$$

أي أن b_2 معنوية إحصائيًا عند مستوى معنوية 5% (وأيضًا 1%) (أي أنه لا يمكن رفض H_1 بأن $b_2 \neq 0$ كون فترة الثقة 95% لكل من (أ)b_1 ، (ب) b_2 في المسألة 7−5 (أ).

(أ) فترة الثقة 95% للمعلمة b_1:

$$b_1 = \hat{b}_1 \pm 2.179 s_{\hat{b}_1} = -0.38 \pm 2.179(0.14) = -0.38 \pm 0.31$$

أي أن b_1 بين -0.69 ،-0.07(أي $-0.69 \leq b_1 \leq -0.07$) بدرجة ثقة 95%.

(ب) فترة الثقة 95% المعلمة b_2:

$$b_2 = \hat{b}_2 \pm 2.179\, s_{\hat{b}_2} = 0.45 \pm 1.179(0.10) = 0.45 \pm 0.22$$

أي أن b_2 بين 0.23، 0.67 (أي $0.23 \le b_2 \le 0.67$) بدرجة ثقة 95%

3. معامل التحديد المتعدِّد:

7−10: بدءًا باستخدام $R^2 = 1 - \sum e_i^2/\sum y_i^2$ اشتق $R^2 = \left(\hat{b}_1 \sum yx_1 + \hat{b}_2 \sum yx_2\right)/\sum y_i^2 \big|$ (إرشاد: ابدأ بتبيان أن $\sum e_i^2 = \sum y_i^2 - \hat{b}_1 \sum yx_1 - \hat{b}_2 \sum yx_2$. يمكن القارئ غير الملم بالتفاضل أن يتخطي هذه المسألة) (دومينيك، 1983 صفحة 178).

$$\sum e_i^2 = \sum e_i \left(y_i - \hat{y}_i\right) = \sum e_i \left(y_1 - \hat{b}_1 x_{1i} - \hat{b}_2 x_{2i}\right)$$
$$= \sum e_i y_i - \hat{b}_1 \sum e_i x_{1i} - \hat{b}_2 \sum e_i x_{2i}$$

ولكن من طريقة OLS وجدنا:

$$\frac{\partial \sum e_i^2}{\partial \sum \hat{b}_1} = -\sum e_i x_{1i} = 0 \quad and \quad \sum e_i x_{1i} = 0$$
$$\frac{\partial \sum e_i^2}{\partial \sum \hat{b}_2} = -\sum e_i x_{2i} = 0 \quad and \quad \sum e_i x_{2i} = 0$$

والتالي:

$$\sum e_i^2 = \sum e_i y_i = \sum \left(y_i - \hat{y}_i\right) y_i = \sum y_i \left(y_i - \hat{b}_1 \sum y_1 x_{1i} - \hat{b}_2 \sum y_i x_{2i}\right)$$
$$= \sum y_i^2 - \hat{b}_1 \sum y_1 x_{1i} - \hat{b}_2 \sum y_i x_{2i}$$

بالتعويض في معادلة R^2 :

$$\mathrm{R}^2 = 1 - \frac{\sum e_i^2}{\sum y_i^2} = 1 - \frac{\sum y_i^2 - \hat{b}_1 \sum y_1 x_{1i} - \hat{b}_2 \sum y_i x_{2i}}{\sum y_i^2} = \frac{\hat{b}_1 \sum y_1 x_{1i} - \hat{b}_2 \sum y_i x_{2i}}{\sum y_i^2}$$

أو بحذف i للتبسيط، نحصل عل (كما في قسم 7−3):

$$\mathrm{R}^2 = \frac{\hat{b}_1 \sum xy_1 + \hat{b}_2 \sum xy_2}{\sum y^2}$$

7−11: أوجد R^2 من معادلة انحدار OLS المقدرة في المسألة 7−5 (أ)، باستخدام: (أ) $R^2 = \sum \hat{y}_i^2/\sum y_i^2$. (ب) $R^2 = 1 - \sum e_i^2/\sum y_i^2$ (ج) $R^2 = \left(\hat{b}_1 \sum yx_1 + \hat{b}_2 \sum yx_2\right)/\sum y_i^2$

(أ) نعرف من المسألة 6−20 أن:

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum e_i^2 \;\; so\ that \;\; \sum \hat{y}_i^2 = \sum y_i^2 - \sum e_i^2$$

حيث $\sum y_i^2 = 40$ (بتربيع وجمع قيم y_i من جدول 7−4) $\sum e_i^2 = 12.2730$ (من جدول 7−5)،

$$\sum \hat{y}_i^2 = 40 - 12.2730 = 27.7270$$

وبالتالي $R^2 = \sum \hat{y}_i^2 / \sum y_i^2 = 27.7270/40 \cong 0.6930 \; or \; 69.32\%$

(ب) باستخدام $\sum e_i^2 = 12.2730$ و $\sum y_i^2 = 40$ نحصل على $R^2 = 1 - \sum e_i^2 / \sum y_i^2 = 1 - 12.2730/40 \cong 0.6932$ أو 69.32% كما في (أ).

(ج) باستخدام $b_1 = 0.38$ و $b_1 = 0.45$ (السابق إيجادها في مسالة 7-5 (أ)، $\sum yx_1 = -28$ و $\sum yx_2 = 38$ (من جدول 7-4)، $\sum y_i^2 = 40$ نحصل على:

$$R^2 = \frac{\hat{b}_1 \sum yx_1 + \hat{b}_2 \sum yx_2}{\sum y_i^2} = \frac{(-0.38)(-28) + (0.45)(38)}{40} \cong \frac{27.74}{40}$$
$$= 0.6935 \; or \; 69.35\%$$

وتختلف قيمة R^2 هذه قليلا عن تلك السابق إيجادها في (أ)، (ب) كنتيجة لأخطاء التقريب.

7-12: من $R^2 = 1 - \left(\sum e_i^2 / \sum y_i^2\right)$ اشتق $\bar{R}^2$ (ب) ما هو المدى لقيم $\bar{R}^2$ (إرشاد بالنسبة لجزء (أ): ابدأ بالتشابه بين $\sum e_i^2$ و تباين $\sum y_i^2$ وتباين y).

(أ) صعوبة R^2 (غير المعدلة) أنها لا تأخذ في الاعتبار درجات الحرية ولكن:

$\mathrm{e} = s^2 = \sum e_i^2 / (n-k)$ حيث: $n - k = df$ وبالتالي: $\sum e_i^2 = s^2(\mathrm{n} - \mathrm{k})$

و $\sum (Y_i - \bar{Y})^2 = \sum y_i^2 = varY(n-1)$

$$R^2 = 1 - \frac{\sum e_i^2}{\sum y_i^2} = 1 - \frac{s^2(n-k)}{var\,Y(n-k)}$$

وعليه: $1 - R^2 = (s^2/var\,Y)(n-k)/(n-1)$ ولكن: $1 - \bar{R}^2 = s^2/var\,Y$ فتكون:

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

وبالحل لإيجاد R^2، نحصل على:

$$\bar{R}^2 = 1 - (1 - \bar{R}^2)\frac{(n-1)}{(n-k)} \qquad (12-7)$$

(ب) عندما $k = 1$ ، $\frac{(n-1)}{(n-k)} = 1$ تكون $R^2 = \bar{R}^2$

عندما $k > 1$ ، $\frac{(n-1)}{(n-k)} > 1$ تكون $R^2 > \bar{R}^2$

عندما تكون n كبيرة، لقيمة معينة k ، تكون $\frac{(n-1)}{(n-k)}$ قريبة من الوحدة، ولن تختلف $\bar{R}^2$ عن R^2 كثيراً.

عندما تكون n صغيرة وتكون k كبيرة بالنسبة إلى n ، فإن $\bar{R}^2$ سوف تكون أصغر كثيرًا من R^2 وقد تكون R^2 سالبة بالرغم من أن $0 \le R^2 \le 1$: (انظر المسائل من 7−29 إلى 7−32.

7−13: (أ) أوجد $\bar{R}^2$ بالنسبة لمعادلة انحدار OLS المقدرة في مسألة 7−5 (أ).

(ب) كيف تقارن $\bar{R}^2$ المحسوبة في (أ) مع $\bar{R}^2$ في مسألة 7−11 (أ)، في مسألة 6−31 (ج)؟.

(أ) باستخدام $R^2 = 0.6932$ السابق إيجادها في المسألة 7−11 (ب)، نحصل على:

$$\bar{R}^2 = 1 - (1 - \bar{R}^2)\frac{(n-1)}{(n-k)} = 1 - (1 - 0.6932)\frac{15-1}{15-3} \cong 0.6410$$

(ب) $R^2 = 0.33$ في حالة الانحدار البسيط، باستخدام نسبة القوة العاملة في الزراعة X_1، كمتغير مستقل وحيد (انظر المسألة 6−31 (ج)). $R^2 = 0.69$ بعد إضافة سنوات التعليم للسكان فوق سن 25 سنة، X_2 كمتغير مستقل ثان. ولكن عندما نأخذ في الاعتبار حقيقة أن إضافة X_2 يقلل درجات الحرية بمقدار 1 في $n - k = 15 - 2 = 13$ في حالة الانحدار البسيط للمتغير Y على X_1، إلى $n - k = 15 - 3 = 12$ في الانحدار المتعدِّد للمتغير Y على X_1 و X_2)، فإن $\bar{R}^2$ تنخفض إلى 0.64

وحقيقة أن b_2 وجدت معنوية إحصائيًا (في المسألة 7−8 (ب)) $R^2 = \bar{R}^2 = 0.33$ في حالة الانحدار البسيط للمتغير Y على X_1 ترتفع إلى $\bar{R}^2 = 0.64$ في حالة الانحدار المتعدِّد للمتغير Y على X_1 و X_2 يبرر الإبقاء على X_2 كمتغير مستقل إضافي في معادلة الانحدار.

7−14: (أ) كيف يمكن إيجاد $\sum e_i^2$ (المطلوبة لإجراء اختبارات المعنوية) بدون إيجاد $\hat{Y}_i$ أولا ؟ (ب) أوجد $\sum e_i^2$ لبيانات جدول 3−7 بدون إيجاد $\hat{Y}_i$ (جدول 7−5).

(أ) باستخدام القيم المقدرة لكمل من b_1 و b_2 وكذلك $\sum yx_1$، $\sum yx_2$ نحصل أولاً على:

$$R^2 = \frac{\hat{b}_1 \sum yx_1 + \hat{b}_2 \sum yx_2}{\sum y^2}$$

وبالتالي $R^2 = 1 - \left(\sum e_i^2 / \sum y_i^2\right)$ فتكون $\sum e_i^2 = (1 - R^2)\sum y_i^2$. وهذه الطريقة لإيجاد $\sum e_i^2$ تتضمن حسابات أقل عن استخدام Y_i (فالحسابات الوحيدة الإضافية بجانب تلك المطلوبة لتقدير b_1 و b_2 هي $\sum y_i^2$).

(ب) من قيمة $R^2 = 0.6935$ السابق إيجادها في المسألة 7−11 (ج) (التي تستخدم فقط القيم المقدرة لكل من b_1 و b_2 السابق إيجادها في مسألة 7−5 (أ) والقيم المحسوبة في جدول 7−4). ومن قيمة $\sum y_i^2 =$ 40 من المسألة $7 - 11$ (أ)، نحصل على:

$$\sum e_i^2 = (1 - R^2)\sum y_i^2 = (1 - 0.6935)(40) = 12.26$$

قارن هذه بقيمة $\sum e_i^2 = 12.2730$ السابق إيجادها في جدول 7 –5. الفرق الصغير في قيمتي $\sum e_i^2$ اللتين حصلنا عليهما باستخدام الطريقتين راجع إلى أخطاء التقريب). لاحظ، أن إيجاد $\sum e_i^2$ بالطريقة السابقة يلغى تمامًا الحاجة إلى جدول 7–5.

4.اختبار المعنوية الإجمالية للانحدار:

7 –15: أُذكر الفرض العدمي والفرض البديل لاختبار معنوية الانحدار ككل. (ب) كيف تختبر المعنوية الكلية للانحدار؟ ما هو منطق هذا الاختبار؟ (ج) أعط صيغة التباين المفسر، التباين غير المفسر أو تباين البواق.

(أ) يشير اختبار المعنوية الإجمالية للانحدار إلى اختبار الفرض أن كل المتغيرات المستقلة لا تساعد على تفسير التغير في المتغير التابع حول وسطه. وبشكل محدّد.

الفرض العدمي هو: $H_0: b_1 = b_2 = \cdots = b_k = 0$ مقابل الفرض البديل.

ليست كل قيم b_i تساوى الصفر: H_i.

(ب) تختبر المعنوية الكلية للانحدار بحساب النسبة F بين التباين المفسر والتباين غير المفسر أو تباين البواقي وتوحى القيمة «المرتفعة» الإحصائية F بعلاقة معنوية بين المتغير التابع والمتغيرات المستقلة، مؤدية إلى رفض الفرض العدمي بأن معاملات كل المتغيرات المفسرة كلها أصفار.

(ج) التباين المفسّر: $\sum \hat{Y}_i^2/(k-1) = RSS/(k-1) = \left(\hat{Y}_i - \bar{Y}\right)^2/(k-1)$ حيث k عدد المعالم المقدرة (انظر قسم 6–4).

والتباين غير المفسر: $\sum e_i^2/(n-k) = ESS/(n-k) = \sum\left(Y_i - \hat{Y}_i\right)^2/(n-k)$

7–16: (أ) أعط صيغة إحصائية أو نسبة F المحسوبة لحالة الانحدار البسيط وللانحدار عند $n = 15$، $k = 3$. (ب) هل يمكن أن تكون F المحسوبة «كبيرة، ومع ذلك فكل المعالم المقدرة ليست معنوية إحصائيًا؟

(أ) $$F_{1.n-2} = \frac{\sum \hat{y}_i^2/1}{\sum e_i^2/(n-2)}$$

حيث تشير رموز دليل F إلى عدد درجات الحرية في البسط والمقام على الترتيب في حالة الانحدار البسيط، $F_{1.n-2} = t_{n-2}^2$ لنفس مستوى الثقة. بالنسبة للانحدار المتعدِّد عند $n = 15$ | $k = 3,$ و $F_{2.12} =$ $\left(\sum \hat{y}_i^2/2\right)/\left(\sum e_i^2/12\right)$.

(ب) من الممكن أن تكون F المحسوبة كبيرة، وليس بين المعالم المحسوبة ما هو معنوي إحصائيًا. وقد يحدث هذا عندما يكون هناك ارتباط مرتفع بين المتغيرات المستقلة بعضها البعض (انظر قسم 9–2). وغالبًا ما

يكون اختبار F ذا فائدة محدودة لأنه من الممكن أن يرفض الفرض العدمي، بصرف النظر عما إذا كان النموذج يشرح «جزءًا كبيرًا» من التغير في Y.

7–17: (أ) أثبت أن:

$$\left[\sum \hat{y}_i^2/(k-1)\right]/\left[\sum e_i^2/(n-k)\right] = [R^2/(k-1)]/[(1-R^2)/(n-k)]$$

(ب) على ضوء نتائج (أ)، ما هي الطريقة البديلة للتعبير عن الفرض لاختبار المعنوية الكلية للانحدار؟

(أ)– $$\frac{\sum \hat{y}_i^2/(k-1)}{\sum e_i^2/(n-k)} = \frac{\sum \hat{y}_i^2}{\sum e_i^2}\frac{n-k}{k-1} = \frac{\sum \hat{y}_i^2/\sum y_i^2}{\sum e_i^2/\sum y_i^2}\frac{n-k}{k-1} = \frac{R^2}{(1-R^2)}\frac{n-k}{k-1} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

(ب) نسبة F ، كاختبار لمعنوية القدرة التفسيرية لكل المتغيرات المستقلة معًا، تعادل تقريبًا اختبار معنوية الإحصائية R^2 فإذا قبل الفرض البديل فإننا نتوقع أن تكون R^2، وبالتالي F «عالية».

7 – 18 اختبر عند مستوى معنوية 5% المعنوية الإجمالية لانحدار OLS المقدر في المسألة 7–5 (أ) $[\sum \hat{y}_i^2/(k-1)]/\left(\sum e_i^2/(n-k)\right)$.

(ب)– $[R^2/(k-1)]/[(1-R^2)/(n-k)]$

(أ) باستخدام: $\sum \hat{y}_i^2 = 27.727$ من المسألة 7–11(أ) ، $\sum e_i^2 = 12.2730$ من جدول 7–5، نحصل على:

$$F_{2.12} = \frac{27.727/2}{12.273/12} \cong 13.59$$

وحيث أن القيمة المحسوبة للنسبة F تفوق القيمة الجدولية F = 3.88 عند مستوى معنوية 5% ودرجات حرية 2، 12 (انظر ملحق 7)، فإننا تقبل الفرض البديل بأنه ليست كل قيم b_i تساوى الصفر عند مستوى معنوية 5%.

(ب) باستخدام $R^2 = 0.6932$ من تمرين 7–11 (ب)، نحصل على:

$$F_{2.12} = \frac{R^2/(k-1)}{(1-R^2)/n-k} = \frac{0.6932/2}{(1-0.6932)/12} \cong 13.54$$

ونقبل الفرض أن R^2 تخطف معنويًا عن الصفر عند مستوى معنوية 5%

5. معاملات الارتباط الجزئـي:

7 – 19 (أ) كيف يمكن إبعاد تأثير X_2 من كل من Y و X_1 عند إيجاد $r_{YX_i.X_{2i}}$؟ (ب) ما هو المدى لقيم معادلات الارتباط الجزئي ؟ (ج) ما هي إشارة معاملات الارتباط الجزئي ؟ (د) ما فائدة معاملات الارتباط الجزئي؟

(أ) لإبعاد تأثير X_2 على Y ، فإننا نوجد انحدار Y على X_2، ونوجد البواقي $e_1 = Y^*$. ولإبعاد تأثير X_2 على X_1 فإننا نوجد انحدار X_1 و X_2 ونوجد البواقي $e_2 = X_1^*$. وفي هذه الحالة فإن Y^*. و X_1^* تمثلان التغير في Y و X_1 على الترتيب، الباقي بدون تفسير بعد إزاحة تفسير X_2 على كل من Y و X_1 وبالتالي، فعامل الارتباط الجزئي ليس إلا معامل الارتباط البسيط بين البواقي Y^* و X_1^* (أى أن، $r_{YX_1.X_2} = r_{Y^*X_1^*}$)

(ب) مدى معاملات الارتباط الجزئي هو من -1 إلى $+1$ (تمامًا كما في حالة معاملات الارتباط البسيط). على سبيل المثال، تشير $r_{YX_1.X_2} = -1$ إلى الحالة عندما توجد علاقة خطية تامة عكسية بين Y و X_1 بعد إزاحة التأثير المشترك للمتغير X_2 على كل من Y و X_1. ولكن، تثير $r_{YX_1.X_2} = 1$ إلى علاقة خطية تامة طردية صافية بين Y و Y_1.. تشير $r_{YX_1.X_2} = 0$ إلى عدم وجود علاقة بين Y و X_1 بعد إزاحة تأثير X_2 على كل من Y و X_1 وكنتيجة، فإنه يمكن حذف X_1 من الانحدار.

(ج) إشارة معامل الانحدار الجزئي في نفس إشارة المعلمة المقدرة المناظرة. فمثلا، بالنسبة لمعادلة الانحدار المقدرة: $\hat{Y} = \hat{b}_0 + \hat{b}_1X_1 + \hat{b}_2X_2 + r_{YX_1.X_2}$ فإن $r_{YX_1.X_2}$ له نفس إشارة $\hat{b}_1$ وكذلك $r_{YX_2.X_1}$ له نفس إشارة في $\hat{b}_2$.

(د) تستخدم معاملات الارتباط الجزئية في تحليل الانحدار المتعدِّد لتحديد الأهمية النسبية لكل متغير مفسر في النموذج. والمتغير المستقل صاحب أعلى معامل ارتباط جزئي مع المتغير التابع يساهم أكثر من المتغيرات المستقلة الأخرى في القدرة التفسيرية للنموذج ويدخل أولاً في تحليل الانحدار المتعدِّد خطوة – بخطوة. ولكن يجب ملاحظة أن معامل الارتباط الجزئي يعطي مقياسًا لترتيب صافى الارتباط وليس مقياسًا لقيمته، فمجموع معاملات الارتباط الجزئي بين المتغير التابع وكل المتغيرات المستقلة لا يساوى 1 بالضرورة.

7–20: بالفنية للانحدار المقدر في المسألة 7–5 (أ)، أوجد (أ) $r_{YX_1.X_2}$، (ب) $r_{YX_2.X_1}$ (ج) هل تساهم X_1 أو X_2 أكثر في القدرة التفسيرية للنموذج ؟

(أ) لإيجاد $r_{YX_1.X_2}$ فإننا نحتاج أولا إلى إيجاد r_{YX_1}، r_{YX_2} و$r_{X_1X_2}$. باستخدام القيم من جدول 7–4، نحصل على:

$$r_{YX_1} = \frac{\sum x_1 y}{\sqrt{\sum x_1^2}\sqrt{\sum y^2}} = \frac{-28}{\sqrt{60}\sqrt{40}} \cong -0.5715$$

(أ)
$$r_{YX_2} = \frac{\sum x_2 y}{\sqrt{\sum x_2^2}\sqrt{\sum y^2}} = \frac{38}{\sqrt{74}\sqrt{40}} \cong 0.6984$$

$$r_{YX_2} = \frac{\sum x_2 x_1}{\sqrt{\sum x_2^2}\sqrt{\sum x_1^2}} = \frac{-12}{\sqrt{74}\sqrt{60}} \cong -0.1801$$

وبالتـــــــــالي:
$$r_{YX_1.X_2}=\frac{r_{YX_1}-r_{YX_2}r_{X_1X_2}}{\sqrt{1-r_{X_1X_2}^2}\sqrt{1-r_{YX_2}^2}}=\frac{(-0.5715)-(0.6984)(-0.1801)}{\sqrt{1-(-0.1801)^2}\sqrt{1-0.6984^2}}\cong -0.6331$$

(ب) باستخدام قيم r_{YX_1}، r_{YX_2}، $r_{X_1.X_2}$ السابق حسابها في (أ)، نحصل على:
$$r_{YX_2.X_1}=\frac{r_{YX_2}-r_{YX_1}r_{X_1X_2}}{\sqrt{1-r_{X_1X_2}^2}\sqrt{1-r_{YX_1}^2}}=\frac{(0.6984)-(-0.5715)(-0.1801)}{\sqrt{1-(-0.1801)^2}\sqrt{1-(-0.5715)^2}}\cong 0.8072$$

(ج) حيث أن: $r_{YX_1.X_2}$ تتجاوز القيمة المطلقة للمعامل فإننا نستنتج أن X_2 تساهم أكثر من Y_1 في القدرة التفسيرية النموذج.

مسائل شاملة:

7-21: يعطى جدول 7-6 الكمية المطلوبة (فرضًا) من سلعة ما، Y وسعرها X_1، ودخل المستهلك، X_2 من عام 1971 إلى 1985. (أ) هيء انحدار OLS لهذه المشاهدات.

جدول 7-6 الكمية المطلوبة من سلعة ما، سعرها، ودخل المستهلك، 1971 – 1980:

السنة	Y	X_1	X_2
1971	40	9	400
1972	45	8	500
1973	50	9	600
1974	55	8	700
1975	60	7	800
1976	70	6	900
1977	65	6	1000
1978	65	8	1100
1979	75	5	1200
1980	75	5	1300
1981	80	5	1400
1982	100	3	1500
1983	90	4	1600
1984	95	3	1700
1985	85	4	1800

(ب) اختبر عند مستوى معنوية 5% المعنوية الإحصائية لمعالم الميل. (ج) أوجد معامل الارتباط المتعدِّد غير المعدل والمعدل. (د) اختبر المعنوية الكلية للانحدار (هـ) أوجد معاملات الارتباط الجزئي وحدّد أي متغير مستقل يساهم أكثر في قدرة النموذج التفسيرية (و) أوجد معامل المرونة السعرية للطلب η_P، والمرونة الداخلية الطلب، η_M، عند المتوسطات. (ز) ضع جميع النتائج في شكل ملخص مع تقريب كل الحسابات إلى 4 علامات عشرية.

(أ) يعطي جدول 7-7 الحسابات اللازمة لتوفيق الانحدار الخطي.

$$\hat{b}_1 = \frac{(\sum x_1 y)(\sum x_2^2) - (\sum x_2 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$
$$= \frac{(-505)(2{,}800{,}000) - (107{,}500)(-11{,}900)}{(60)(2{,}800{,}000) - (-11{,}900)^2} \cong 5.1061$$

$$\hat{b}_2 = \frac{(\sum x_2 y)(\sum x_1^2) - (\sum x_1 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - \left(\sum x_1\, x_2\right)^2} = \frac{(107{,}500)(60) - (-505)(-11{,}900)}{(60)(2{,}800{,}000) - (-11{,}900)^2}$$
$$\cong 0.1607$$

$$\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X}_1 - \hat{b}_2 \bar{X}_2 = 70 - (-5.1061)(6) - (0.0167)(1{,}100) \cong 82.2666$$
$$\hat{Y} = 82.2666 - 5.1061 X_1 + 0.0167 X_2$$

(ب) ويمكننا إيجاد $\sum e_i^2$ بحساب R^2 أولًا من جدول 7-7:

$$R^2 = \frac{\hat{b}_1 \sum y x_1 + \hat{b}_2 \sum y x_2}{\sum y^2} = \frac{(-5.1061)(-505) + (0.0167)(107{,}500)}{4{,}600} \cong 0.9508$$

ولكن:
$$R^2 = 1 - \frac{\sum e^2}{\sum y^2}$$

وبالتالي:
$$\sum e^2 = (1 - R^2)\sum y^2 = (1 - 0.9508)600 \cong 226.32$$

$$s_{\hat{b}_1}^2 = \frac{\sum e_i^2}{n-k} \frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2} = \frac{226.32}{15-3} \frac{2{,}800{,}000}{(60)(2{,}800{,}000) - (-11{,}900)^2}$$
$$\cong 2.0011 \quad and \qquad s_{\hat{b}_1^2} \cong 1.4146$$

$$s_{\hat{b}_2}^2 = \frac{\sum e_i^2}{n-k} \frac{\sum x_i}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2} = \frac{226.32}{15.3} \frac{60}{(60)(2{,}800{,}000) - (-11{,}900)^2}$$
$$\cong 0.00004 \quad and \quad s_{\hat{b}_2} \cong 0.0065$$

$$t_1 = \frac{\hat{b}_1}{s_{\hat{b}_1}} = \frac{-5.1061}{1.4146} \cong -3.6096 \quad and \quad t_2 = \frac{\hat{b}_2}{s_{\hat{b}_2}} = \frac{0.0167}{0.0065} \cong 2.5692$$

وبالتالي، فإن كلا من $\hat{b}_1$ و $\hat{b}_2$ معنوية إحصائيًا عند مستوى معنوية 5%

(ج) $R^2 = 0.9508$ (من (ب))، وعليه:

$$\bar{R}^2 = 1-(1-R^2)\frac{n-1}{n-k} = 1-(1-0.9508)\frac{15-1}{15-3} \cong 0.9426$$

د)–

$$F_{k-1,n-k} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{0.9508/(3-1)}{(1-0.9508)/(15-3)} \cong 115.9512$$

وبالتالي، فإن R^2 تختلف معنويًا من الصفر عند مستوى معنوية 5%

السنة	Y	X_1	X_2	y	x_1	x_2	yx_1	yx_2	x_1x_2	x_1^2	x_2^2	y^2
1971	40	9	400	−30	3	−700	−90	21,000	−2,100	9	490,000	900
1972	45	8	500	−25	2	−600	−50	15,000	−1,200	4	260,000	625
1973	50	9	600	−20	3	−500	−60	10,000	−1,500	9	250,000	400
1974	55	8	700	−15	2	−400	−30	6,000	−800	4	160,000	225
1975	60	7	800	−10	1	−300	−10	3,000	−300	1	90,000	100
1976	70	6	900	0	0	−200	0	0	0	0	40,000	0
1977	65	6	1,000	−5	0	−100	0	500	0	9	10,000	25
1978	65	8	1,100	−5	2	0	−10	0	0	4	0	25
1979	57	5	1,200	5	−1	100	−5	500	−100	1	10,000	25
1980	57	5	1,300	5	−1	200	−5	1,000	−200	1	40,000	25
1981	080	5	1,400	10	−1	300	−10	3,000	−300	1	90,000	100
1982	100	3	1,500	30	−3	400	−90	12,000	−1,200	9	160,000	900
1983	90	4	1,600	20	−2	500	−40	10,000	−1000	4	250,000	400
1984	95	3	1,700	25	−3	600	−75	15,000	−1800	9	360,000	625
1985	85	4	1,800	15	−2	700	−30	10,500	−1400	4	490,000	225
0	$\sum Y = 1,05$ $\bar{Y} = 75$	$\sum X_1 = 90$ $\bar{X}_1 = 6$	$\sum X_2 = 16,50$ $\bar{X}_2 = 1,100$	$\sum y = 0$	$\sum x_1 = 0$	$\sum x_2 = 0$	$\sum yx_1 = -50$	$\sum yx_2 = 107,50$	$\sum x_1x_2 = 11,90$	$\sum x_1^2 = 60$	$\sum x_2^2 = 2,800,00$	$\sum y^2 = 4,60$

(هـ) لإيجاد $r_{YX_1.X_2}$ و $r_{YX_2.X_1}$، فيجب أولاً إيجاد (من جدول 7–7).

$$r_{YX_1} = \frac{\sum x_1 y}{\sqrt{\sum x_1^2}\sqrt{\sum y^2}} = \frac{-505}{\sqrt{60}\sqrt{4,600}} \cong 0.9613$$

$$r_{YX_2} = \frac{\sum x_2 y}{\sqrt{\sum x_2^2}\sqrt{\sum y^2}} = \frac{107,500}{\sqrt{2,800,000}\sqrt{4,600}} \cong 0.9472$$

$$r_{X_1X_2} = \frac{\sum x_1 x_2}{\sqrt{\sum x_2^2}\sqrt{\sum x_1^2}} = \frac{-11,900}{\sqrt{2,800,000}\sqrt{60}} \cong 0.9181$$

$$r_{YX_1.X_2} = \frac{r_{YX_1} - r_{YX_2}r_{X_1X_2}}{\sqrt{1-r_{X_1X_2}^2}\sqrt{1-r_{YX_2}^2}} = \frac{(-0.9613)-(0.9472)(-0.9181)}{\sqrt{1-(-0.9181)^2}\sqrt{1-(-0.9472)^2}} \cong -0.7213$$

$$r_{YX_2.X_1}=\frac{r_{YX_2}-r_{YX_1}r_{X_1X_2}}{\sqrt{1-r_{X_1X_2}^2}\sqrt{1-r_{YX_1}^2}}=\frac{(0.9472)-(0.9613)(-0.9181)}{\sqrt{1-(-0.9181)^2}\sqrt{1-(-0.9613)^2}}\cong 0.5919$$

وبالتالي فإن X_1 تساهم أكثر من X_2 في القدرة التفسيرية للنموذج.

$$\eta_P=\hat{b}_1\frac{\bar{X}_1}{\bar{Y}}=-5.1061\frac{6}{70}\cong 0.4377$$

(و)

$$\eta_M=\hat{b}_2\frac{\bar{X}_2}{\bar{Y}}=0.0167\frac{1{,}100}{70}\cong 0.2624$$

(ز)

$$\hat{Y}_1=82.2666-5.1061X_1+0.0167X_2$$

$$value\quad(-3.096)\quad(2.5692)$$

$$R^2=0.9508\quad \bar{R}^2=0.9426\quad ,F_{2.12}=155.9512$$

$$r_{YX_1.X_2}=0.7023\quad ,\quad r_{YX_2..X_1}=0.8434$$

$$n_P=-0.4377\qquad n_M=0.2624$$

22–7: جدول 8–7 هو نفس جدول 6–7 فيما عدا أنه يشمل سعر سلعة بديلة بالدولار، X_3 كمتغير مستقل ثالث. جدول 7–9 يعطي صورة من مخرجات الكمبيوتر للانحدار الخطي للمتغير Y على X_1، X_2 ، X_3 باستخدام SPSS (Statistical Package for the Social Sciences أكثر برامج الكمبيوتر شيوعًا.

جدول 8–7 الكمية المطلوبة، السعر، دخل المستهلك، وسلعة بديلة

السنة	Y	X_1	X_2	X_3
1971	40	9	400	10
1972	45	8	500	14
1973	50	9	600	12
1974	55	8	700	13
1974	60	7	800	11
1976	70	6	900	15
1977	65	6	1,000	16
1978	65	8	1,100	17
1979	75	5	1,200	22
1980	75	5	1,300	19
1981	80	5	1,400	20
1982	100	3	1,500	23
1983	90	4	1,600	18
1984	95	3	1,700	24
1985	85	4	1,800	21

أجب عن الأسئلة التالية من خلال فحص مخرجات الكمبيوتر في جدول 9-7. (أ) اكتب معادلة انحدار OLS مع قيم t, R^2, $\bar{R}^2$ نسبة F مع درجات الحرية، الخطأ المعياري للانحدار، ومجموع مربعات البواقي، وفسر النتائج. (ب) كيف أجري برنامج SPSS ؟

(أ) من صفحة 4 من مخرجات الكمبيوتر في جدول 9-7، نحصل على:

$$t_1 = \frac{\hat{b}_1}{s_{\hat{b}_1}} = \frac{-4.9281}{1.6111} = -3.05, \quad t_2 = \frac{\hat{b}_2}{s_{\hat{b}_2}} = \frac{0.0159}{0.0074} = 2.149, t_1 = \frac{\hat{b}_3}{s_{\hat{b}_3}} = \frac{0.1748}{0.6367} = 0.275$$

وعليه:

$$\begin{gathered} Y = 79.1063 - 4.9281X_1 + 0.0159X_2 + 0.1748X_3 \\ (-3.059) \quad (2.149) \quad (0.275) \\ R^2 = 0.95 \ , \quad \bar{R}^2 = 0.94 \ , \qquad F_{3.11} = 71.13 \\ s = 4.53 \ , \ SEE = 225.49 \end{gathered}$$

إشارات $\hat{b}_1$، $\hat{b}_2$، $\hat{b}_3$ تتفق مع ما هو متوقع طبقًا للنظرية الاقتصادية. ولكن $\hat{b}_1$ فقط معنوية إحصائيًا عند مستوى معنوية 5%، بينما كادت $\hat{b}_2$ أن تكون معنوية. من نسبة F، تنبّه إلى أن R^2 (والانحدار ككل) معنوى إحصائيًا عند مستوى 5%. يجب ملاحظة أنه في صفحة 4 من جدول 7-9، R المتعدِّدة هي الحذر التربيعي للمقدار R تربيع وبالتالي تشير إلى معامل الارتباط المتعدِّد r.

جدول 9-7 مخرجات الكمبيوتر SPSS لانحدار الطالب

وتشير بيتا إلى المعاملات المعيارية، أو المعاملات المقدرة مضروبة في نسبة الانحراف المعياري لمتغير مستقل معين إلى الانحراف المعياري المتغير التابع. قيمة F لكل معامل ليست إلا مربع قيمة r، لكل معامل.

(ب) الخطوة الأولى لإجراء برنامج كمبيوتر هي «الدخول» من إحدى محطات الكمبيوتر ويحتاج هذا إلى رقم حساب و «وكلمة مرور» (والتي تحصل عليها من مركز الكمبيوتر في جامعتك، أو شركتك). لاستخدام SPSS فإنك تستخدم أمرًا، يمدك به أيضًا مركز الكمبيوتر. الخطوات التالية موجودة في صفحة 1 (السطور من 5 إلى 15) من مخرجات الكمبيوتر في جدول 9-7. وهذه في معظمها لا تحتاج إلى شرح. في السطر الرابع INPUT (FORMAT) تشير FREEFIELD إلى الطريقة التي يتغذى بها الكمبيوتر بالبيانات، حيث تدخل قيمة كل متغير في التسلسل المشار إليه في سطر 3 (أي $75ect$ 24 1,700 3 95 12 600 9 50) في سطر 7، تشير الأقواس إلى أن المطلوب هو انحدار متعدِّد عادي، بينما تطلب RESID = 0 حساب البواقي المعيارية وقيم التنبؤ للمتغير Y. ويطلب OPITON 11,12, على سطر 8 كتابة البواق المعيارية وقيم التنبؤ للمتغير Y (المعطاة في صفحة 7 من مخرجات الكمبيوتر). وتطلب STATISTICS 1 to 8 في سطر 9 كتابة المتوسط،

الانحراف المعياري وعدد الحالات الصحيحة (السنوات) لكل متغير (المعطاة في صفحة 2 من مخرجات الكمبيوتر؛ مصفوفة الارتباط البسيط (المعطاة في صفحة 3) ؛ كل الإحصائيات في صفحة4 (السابق مناقشتها في (أ))، وإحصائية دير بين واتسون. ارجع إلى دليل SPSS في مركز الكمبيوتر، وأخيرًا، فإن الجدول الموجز في صفحة 5 (ويتضمن كل خرجات الكمبيوتر بشكل أوتوماتيكي (يعالج المتغيرات كما لو كانت أدخلت الكمبيوتر واحدًا وراء الآخر وليس معًا.

6.مسائل إضافية: النموذج الخطي ذو المتغيرات الثلاثة:

7-23: جدول 7-10 امتداد لجدول 6-12 ويعطي مشاهدات عن Y، X_1، X_2. أوجد معادلة انحدار OLS للمتغير Y علي X_1 و X_2.

الإجابة: $\hat{Y}_1 = 4.76 + 5.29X_{1i} + 2.13X_{2i}$

جدول 7-10: مشاهدات عن Y، X_1 و X_2.

n	1	2	3	4	5	6	7	8	9	10
Y	20	28	40	45	37	52	54	43	65	56
X_1	2	3	5	4	3	5	7	6	7	8
X_2	5	6	6	5	5	7	6	6	7	7

7-24 بالرجوع إلى معادلة انحدار OLS المقدرة في المسألة 7-23 فسر (أ)$\hat{b}_0$ و، (ب) $\hat{b}_1$، (ج) $\hat{b}_2$.

الإجابة:

(أ) $\hat{b}_0 = 4,76$ هي الثابت أو مقطع Y، $\hat{Y}_i = \hat{b}_0 = 4.76$ ، عند $X_{i1} = X_{2i} = 0$

(ب) $b_1 = 5.29$ تشير إلى أن زيادة قدرها الوحدة في X_1 (مع تثبيت X_2) ينتج عنها زيادة في $\hat{Y}_i$ قدرها 5.29 وحدة.

(ج) $b_2 = 2.13$، تشير إلى أن زيادة قدرها الوحدة في X_2 (مع تثبيت X_1) ينتج عنها زيادة في $\hat{Y}_i$ قدرها 2.13 وحدة.

1.6 اختبارات معنوية تقديرات المعالم:

7-25: بالرجوع إلى بيانات جدول 7-10، أوجد (أ) S^2، (ب) $s^2_{\hat{b}_1}$ و $s_{\hat{b}_1}$ (ج) $s^2_{\hat{b}_2}$ و $s_{\hat{b}_2}$.

الإجابة: (أ)$S^2 = 50$ (ب) $s_{\hat{b}_1} \cong 1.78$ and $s^2_{\hat{b}_1} \cong 3.16$ (ج) $s^2_{\hat{b}_2} \cong 18.95$ و $s_{\hat{b}_2} \cong 4.35$

7-26 اختبر عند مستوى معنوية 5% كلا من (أ) b_1، (ب) b_2 في المسألة 7-23.

الإجابة: (أ) b_1 معنوية إحصائيًا عند مستوى 5%.

(ب)b_2 ليست معنوية إحصائيًا عند مستوى 5%

7–27 كون فترة الثقة 95% لكل من (أ) b_1، (ب) b_2 في المسألة 7–23.

الإجابة: (أ) $1.08 \leq b_1 \leq 9.50$ (ب) $-8.16 \leq b_2 \leq 12.42$

2.6 معامل التحديد المتعدِّد:

7–28 بالنسبة لانحدار OLS المقدرة في المسألة 7–23، أوجد (أ) R^2، (ب) $\bar{R}^2$ (ج) هل يجب أن تدخل X_2 في الانحدار؟

الإجابة: (أ) $R^2 \cong 0.79$ (باستخدام $R^2 = 1 - \left(\sum e_i^2 / \sum y_i^2\right)$) (ب) $\bar{R}^2 \cong 0.73$ (ج) حيث أن b_2 وجدت غير معنوية إحصائيًا (في المسألة 7–26 (ب))، R^2 نقصت من $R^2 = \bar{R}^2 = 0.77$ عندما كانت X_1 المتغير المستقل الوحيد (انظر المسألة 6–41 (أ)) إلى $\bar{R}^2 = 0.73$ (أعلاه)، فإن X_2 يجب ألا تدخل الانحدار.

7–29 بالنسبة إلى $R^2 = 0.60$، $n = 10$ ، $K = 1$ أوجد $\bar{R}^2$

الإجابة: $\bar{R}^2 = 0.60$

7–30 بالنسبة إلى $R^2 = 0.60$، $n = 10$ ، $K = 2$ ، أوجد $\bar{R}^2$

الإجابة: $\bar{R}^2 = 0.55$

7–31 بالنسبة إلى $\bar{R}^2 = 0.60$، $K = 2$ (كما في المسألة 7–30) ولكن $n = 100$، أوجد $\bar{R}^2$

الإجابة: أوجد $\bar{R}^2 = 0.569$

7–32 بالنسبة إلى $R^2 = 0.40$، $n = 10$ ، $K = 5$ أوجد $\bar{R}^2$.

الإجابة: $\bar{R}^2 = -0.08$ (ولكنها تفسر على أنها تساوى الصفر).

3.6 اختبار المعنوية الكلية للانحدار:

7–33 من انحدار OLS المقدر في المسألة 7–23، أوجد (أ) التباين المفسر، (ب) التباين غير المفسر أو تباين البواقي (ج) نسبة أو إحصائية F

الإجابة: (أ) $\sum \hat{y}^2/(k-1) \cong 649$ (ب) $\sum e^2/(n-k) = 50$ (ج) $F_{2,7} = 12.98$

7–34 اختبر المعنوية الإجمالية لانحدار OLS المقدر في المسألة 7–33 عند (أ) مستوى 5% (ب) مستوى 1%.

الإجابة: (أ) حيث أن نسبة F المحسوبة (12.98) تتجاوز قيمة F النظرية أو الجدولية (4,74) عند$0 = \alpha$ ودرجات حرية 2، 7، فإننا تقبل الفرض بأن معالم انحدار OLS المقدرة هي معًا معنوية عند مستوى 5%.

(ب) حيث أن قيمة F الجدولية عند مستوى $\alpha = 0.01$ هي $F = 9.55$ ، يقبل الفرض البديل عند مستوى معنوية 1% أيضًا.

4.6 معاملات الارتباط الجزئية:

7−35 بالنسبة لانحدار OLS المقدر في المسألة 7−23، أوجد (أ) r_{YX_1,X_2}، (ب) r_{YX_2,X_1} (ج) أي ستغير مستقل يساهم أكثر في القدرة التسييرية للنموذج ؟

الإجابة: (أ) 0.74 (ب) $r_{YX_2,X_1} = 0.18$ (ج) X_1

7. مسألة شاملة:

7−36 جدول 7−11 امتداد للجدول 6−13 ويعطى بيانات عينة عشوائية من 12 من الأسر عن عدد الأطفال في الأسرة Y، وعدد الأطفال الذين قالوا إنهم كانوا يرغبون في إنجابهم وقت الزواج X_1، وعدد سنوات تعليم الزوجة، X_2. (أ) أوجد معادلة انحدار OLS للمتغير Y على X_1 و X_2. (ب) احسب قيم t، واختبر عند مستوى 5% المعنوية الإحصائية لمعالم الميل. (ج) أوجد معامل الارتباط المتعدِّد غير المعدل والمعدل (د) اختبر المعنوية الإجمالية للانحدار (هـ) أوجد معاملات الارتباط الجزئية وحدد أي المتغيرات المستقلة يساهم أكثر في القدرة التفسيرية للنموذج. قم بجميع الحسابات إلى رقمين عشريين.

جدول 7−11 عدد الأطفال في الأسرة وعدد الأطفال المرغوب فيهم وتعليم الزوجة

الأسرة	1	2	3	4	5	6	7	8	9	10	11	12
Y	4	3	0	4	4	3	0	4	3	1	3	1
X_1	3	3	0	2	2	3	0	3	2	1	3	2
X_2	12	14	18	10	10	14	18	12	15	16	14	15

الإجابــة: (أ) $\hat{Y} = 6.90 + 0.53X_1 - 0.39X_2$ (ب) حيث أن $t_1 = 3.12$ و $t_2 = -5.57$ فإن كلا من $\hat{b}_1$ $\hat{b}_2$ معنوية إحصائيًا عند مستوى 5% (ج) $R^2 = 0.95$ و $\bar{R}^2 = 0.90$ (د) حيث أن $F_{2,9} = 51.31$ فإن R^2 معنوية إحصائيًا عند مستوى 5% (هـ) $r_{YX_2.X_1} = 0.87$ و $r_{YX_1.X_2} = 0.71$ وبالتالي فإن X_2 تساهم أكثر من X_1 في القدرة التفسيرية النموذج.

المحور الخامس:

المشاكل القياسية- الارتباط الذاتي للأخطاء، عدم ثبات تباين الأخطاء، التوزيع غير الطبيعي للأخطاء

1.مشكلة عدم ثبات تجانس التباين Hetroscadasticity problem

سبق أن تم الإشارة إلى فروض المتغير العشوائي على أن تباين المتغير يجب أن يكون متجانسًا في المجتمعات الفرعية حتى يكون تباينه في المجتمع الإحصائي ككل متجانسًا وأي خلاف حول هذا المنطق سيؤدي إلى خرق للفرضية المنوه عنها سلفا، وتحصل مشكلة تسمى بمشكلة عدم ثبات تباين التجانس أي أن:

$$E(UiUj) \neq 0 \dots\dots i \neq j \dots (i,j = 1,2,3 \dots\dots n)$$
$$E(Ui^2) = oui^2$$
$$ou1^2 \neq ou2^2 \neq ou3^2 \neq \cdots \dots\dots \neq oun^2$$

وبالتالي فإن التشوه في التقدير حاصل لامحالة ولا يمكن الاعتداد بنتائج التقديرات لاتخاذ قرارات اقتصادية أو مصيرية أو القيام بالتنبؤ المستقبلي لعدم صحة هذه التقديرات.

إن هذه المشكلة غالبا ما تحدث في النماذج التي تعتمد على البيانات المقطعية cross–section وفي بيانات ذات نوع يكون فيه التفاوت كبيرا، مما يؤدي إلى تفاوت تباين الخطأ العشوائي، بحيث تارة يكون كبيرا وتارة أخرى يصبح صغيراً، ويمكن أن تحدث هذه في بيانات الاتفاقات الأسرية أو ما يسمى انفاق العوائل إلى السلع الغذائية، إذ ينصب كل دخول العوائل المتدنية الدخل على السلع الضرورية بينما تختلف الأسر ذات الدخل المتوسط والعالي عن تواجهات الاتفاق عن الأسر ذات الدخل المنخفض، مما يجعل هذه البيانات متفاوتة بشكل كبير وهذه الحالة ستؤدي إلى ما يأتي (عدنان داوي، 2010 صفحة 107):

$$E(Ui^2) \neq oui^2$$
$$E(UiUj) \neq 0$$
$$E(U'U) \neq oui^2In$$

مما يشكل خرقا واضحا كما أسلفنا للفرضية، وبالتالي لا يمكن من تقديرها بالطريقة المعهودة طريقة المربعات الاعتيادية لأنه لا يكون تقديرا غير متحيز وبالتالي يستوجب تقديرها بطرق أخرى، إلا أن من البديهي أن نعرف أن البيانات هي أساس هذه المشكلة، لذا يستوجب تنقيتها من القيم الشاذة والتي قد تسبب مشكلة أخرى عندما يتم تقدير النموذج بالطرق الأخرى ويمكن استخدام طريقة المربعات الصغرى المعممة GLS) Generalized least squares method)

وكما يأتي:

يتم تنقية البيانات باستخدام ما يعرف بـ: معكوس مصفوفة اوميغا Ω^{-1} والتي يمكن الحصول عليها من المصفوفة المعروفة بالرمز P وتساوي:

$$P=\begin{bmatrix}\frac{1}{2\sqrt{\lambda 1}} & 0 & 0\\ 0 & \frac{1}{2\sqrt{\lambda 2}} & 0\\ 0 & 0 & \frac{1}{2\sqrt{\lambda 3}}\end{bmatrix}$$

ومن ثم نحصل على معكوسها وكما يأتي:

$$P-1=\begin{bmatrix}\frac{1}{2\sqrt{\lambda 1}} & 0 & 0\\ 0 & \frac{1}{2\sqrt{\lambda 2}} & 0\\ 0 & 0 & \frac{1}{2\sqrt{\lambda 3}}\end{bmatrix}$$

ومن ثم نقوم بضرب المصفوفة بمقلوبها لتصبح كما يأتي :

$$P'P=\begin{bmatrix}\frac{1}{\lambda 1}\ldots 0\ldots..0\ldots\ldots\ldots.0\\ 0\ldots..\frac{1}{\lambda 1}\ldots.0\ldots\ldots\ldots 0\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots..\\ 0\ldots\ldots.0\ldots 0\ldots\ldots\frac{1}{\lambda 1}\end{bmatrix}$$

وهذه المصفوفة تعرف بمصفوفة أوميغا Ω وإن معكوسها يصبح كما يأتي:

$$(P'P)^{-1}=P\quad P^{-1}\begin{bmatrix}\lambda 1\ldots\ldots.0\ldots\ldots..0\ldots\ldots\ldots\ldots 0\\ 0\ldots..\lambda 2\ldots\ldots 0\ldots\ldots\ldots\ldots\ldots 0\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..0\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..0\\ 0\ldots\ldots 0\ldots.0\ldots\ldots\ldots\ldots\ldots..\lambda n\end{bmatrix}=\Omega^{-1}$$

وعليه أصبح من الممكن أن نشتق صيغة طريقة المربعات الصغرى المعممة كما يأتي :-

$$y=XB+U$$
$$U=-XB$$

نضرب طرفي المعادلة P^{-1} : $\quad P^{-1}U=P^{-1}Y-P^{-1}XB$

تربع المعادلة :

$$(P^{-1}U)(P^{-1}U)=(P^{-1}Y-P^{-1}XB)(P^{-1}Y-P^{-1}XB)$$
$$P^{-1}U'UP^{-1}=(Y'P^{-1}-X'B'P^{-1\prime})(YP^{-1}-XBP^{-1})$$
$$P^{-1}U'UP^{-1}=Y'P^{-1\prime}YP^{-1}-Y'^{P^{-1\prime}}-XBP^{-1}-X'B'P^{-1\prime}YP^{-1}+X'B'P^{-1\prime}XBP^{-1}$$
$$Y'P^{-1\prime}XBP^{-1}=X'B'P^{-1\prime}YP^{-1}\ldots P^{-1}P^{-1}=\Omega^{-1}.$$
$$(U^{-1}\Omega^{-1}U)=Y'\Omega^{-1}Y-2X'B'\Omega^{-1}Y+X'B'\Omega^{-1}XB$$

نأخذ الاشتقاق الجزئي الأول لمربعات البواقي بالنسبة إلى مقلوب موجه المعالم:

$$\frac{\partial(U^{-1}\Omega^{-1}U)}{\partial B'} = -2X'\Omega^{-1}Y + 2X'\Omega^{-1}XB = 0$$

نقسم طرفي المعادلة على 2:

$$\frac{\partial(U^{-1}\Omega^{-1}U)}{\partial B'} = -X'\Omega^{-1}Y + (X'\Omega^{-1}X)B = 0$$

وهي صيغة تقدير طريقة المربعات الصغرى المعممة.

$$\therefore B = \frac{X'\Omega^{-1}Y}{(X'\Omega^{-1}X)}$$

$$B = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$$

وهي صيغة تقدير طريقة المربعات الصغرى المعمة.

أما صيغة التباين والتباين المشترك، فمن صيغة التباين السالفة الذكر في طريقة المربعات الصغرى الاعتيادية والتي تأخذ الشكل الآتي :

$$vae - cov(b) = E[b - E(b)]^2$$

$$\therefore E(b) = B$$

ومن صيغة طريقة المربعات المعممة الآنفة الذكر:

$$[b] = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$$

$$\because Y = XB + U$$

$$\therefore [b] = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}(XB + U)$$

$$[b] = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\mathrm{XB} + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}U$$

$$\because (X'\Omega^{-1}X)^{-1} = X'\Omega^{-1}X = 1$$

$$\therefore [b] = IB + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}U$$

$$[b] = B + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}U$$

$$[b - B] = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}U$$

$$\because var - cov[b] = E[b - B]^2$$

$$\because var - cov[b] = E[((X'\Omega^{-1}X)^{-1}X'\Omega^{-1}U)/(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}U]$$

$$var - cov[b] = [((X'\Omega^{-1}X)^{-1}X'\Omega^{-1}U)/(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}U]$$

$$\therefore (X'\Omega^{-1}X)^{-1} = X'\Omega^{-1}X = I \,....\, U(U = ou^2$$

$$\therefore var - cov[b] = ou^2(X'\Omega^{-1}X)^{-1}$$

وعند مقارنة تباينات الطريقة المعممة بالطريقة الاعتيادية لاستبيان الطريقة الأكفأ في التقدير ووفقا لما يأتي:

$$\frac{var(\ddot{b})inGLS}{var(\ddot{b})inOLS} = 1\,, > 1, < 1$$

فاذا كانت النتيجة تساوي 1 فيعني أن الطريقتين متساويتين في الكفاءة، أما إذا كانت النتيجة أقل من الواحد الصحيح فيعني هذا أن طريقة المربعات الصغرى المعممة أكفأ من الطريقة الاعتيادية، وأما إذا كانت النتيجة أكبر من الواحد الصحيح فالطريقة الاعتيادية أكفأ من الطريقة المعممة.

2.اختبار وجود مشكلة عدم ثبات تجانس التباين :

هناك طرق عدة للكشف عن وجود أو عدم وجود المشكلة في النموذج المقدر ومن هذه الاختبارات وأكثرها شهرة:

1.2 اختبار معامل ارتباط الرتب لسبيرمان Spearman Rank correlation coefficient

إذ يعتبر هذا الاختبار من أفضل الاختبارات من حيث السهولة والشمولية لكل النماذج المقدرة البسيطة والمتعدِّدة . ويمكن تطبيق هذا الاختبار كما يأتي:

- يتم تقدير نموذج الانحدار بالطريقة التي أشرتا إليها وهي طريقة المربعات الصغرى الاعتيادية.
- يحتسب قيم التغيير التابع المقدر $\dot{Y}$ بواسطة إدخال قيم المتغير المستقل للحصول على هذا العمود.
- يطرح العمود المقدر من عمود القيم الحقيقية لـ $\dot{Y}$ لاحتساب قيم المتغير العشوائي $UI = YI - \dot{Y}I$
- ترتب قيم المتغير العشوائي أو البواقي ترتيبا إما تصاعديا أو تنازليا مع المتغير المستقل .
- نحتسب الفرق ما بين العمودين وذلك بطرح عمود من العمود الآخر ولا يشترط أن يكون عمودا معينا ويوضع الفرق في عمود يرمز له Di .
- تربع القيم في عمود Di وتوضع في عمود آخر Di^2 وتجمع القيم $\sum \mathrm{Di}^2$ وهو مجموع مربعات الفروق.

للحصول على المعامل يتم من خلال الصيغة الآتية:

$$rex = 1 - \frac{3 \sum \mathrm{Di}^2}{n(n^2 - 1)}$$

$$rex = Spearman\ coefficient$$

$$n = sample\ size$$

والنتيجة المرتقبة من تطبيق الصيغة أعلاه تنحصر قيمتها ما بين الصفر والواحد الصحيح ، فكلما اقتربت قيمة معامل ارتباط الرتب من الواحد الصحيح دل ذلك على وجود علاقة قوية ما بين قيم البواقي وقيم المتغير المستقل وبالتالي إلى وجود مشكلة عدم ثبات تجانس التباين والعكس صحيح .

2.2 اختبار بارتليت Bartlett test:

يقوم هذا الاختبار على احتساب تباين الخطأ العشوائي auit فعندما oui^2 ، فعندما يكون هذا التباين ثابتا في المجتمعات الفرعية من المجتمع الإحصائي برمته يعني ذلك ثبات تجانس التباين وبطبيعة الحال في الإحصاء فإنها تخضع للفرضيات التي تم ذكرها سلفا؛ وهما فريضة العدم والفرضية البديلة، إذ أن فرضية العدم HO تدل على ثبات تجانس التباين والفرضية البديلة $H1$ تدل على عدم ثبات تجانس التباين أي أن:

Ho: oul2 = ou22 = ou3.. Hl: oulou22 ou32..

$$H0:\ \mathrm{ou}1^2 = ou2^2 = ou\ 3^2 \ldots\ldots\ldots\ldots = ou\ n^2$$
$$H1:\ \mathrm{ou}I^2 \neq ou2^2 \neq ou\ 3^2 \ldots\ldots\ldots\ldots \neq ou\ n^2$$

ويكون الاختبار ذا جودة عالية عندما تكون البيانات يتوافر فيها الآتي:

أ – عندما تكون قيم المتغير المستقل مختلفة بعضها عن البعض الآخر.

ب – عندما تكون هناك مستويات رئيسة وأخرى ثانوية مثال ذلك بيانات الأسر ولغرض إجراء الاختبار يتبع مابلي:

– تقسم البيانات إلى أجزاء ولتكن m من الأجزاء، وكل جزء يحتوي على عدد من العينات N.

– يقدّر تباين الخطأ العشوائي $\mathrm{ou}i^2$ لكل عينة.

إيجاد قيم كل من L ، Q والتي تأخذ الصيغ الآتية (عدنان داوي، 2010 صفحة 112) :–

$$Q = N\log\left(\frac{\sum_{i=1}^{m} nioui^2}{N}\right) - \sum_{i=1}^{m} ni\log oui^2$$
$$L = 1 + \frac{1}{3(m-1)}\left(\sum_{i=1}^{m}\frac{1}{ni} - \frac{1}{N}\right)$$

N عدد العينات الكلي (جميع العينات في كل أجزاء العينة المقسمة).

ni حجم العينة المجزأة.

Log = اللوغارتيم الطبيعي أساس 10.

ج– يحتسب اختبار بارتليت كما يأتي: Q/I.

والقيمة الناتجة تتقارب مع توزيع (كاي سكويرز $\chi^2 n - 1$) أي تختبر بحسب جدول كاي سكويرز بدرجة حرية ($\mathrm{m} - 1$) ومستوى معنوية معين فإذا كانت القيمة المحسوبة لاختبار بارتليت أقل أو تساوي الجدولية عند ذلك المستوى من المعنوية تقبل فرضية العدم بوجود ثبات تجانس التباين ، أما إذا كانت القيمة المحتسبة أكبر من

الجدولية فيعني ذلك أن تباين الخطأ المحتسب للعينات الجزئية غير متجانس يدل على أن تلك العينات تعود إلى مجتمع واحد وليس إلى مجتمعات فرعية .

3.2 اختبار كولد فيلد كوانت Goldfeld& Quandt test

يقوم هذا الاختبار على الآتي:–

A–ترتب قيم المتغير المستقل Xi ترتيبا تصاعديا

B– يتم حذف القيم الوسطية من قيم المتغير المستقل التي تم ترتيبها بواقع 8–9 عينات في حالة كون حجم العينة تبلغ نحو 30 عينة وبواقع 16–18 عينة في حالة حجم العينة الكلي يبلغ نحو 60 عينة.

C–تقسم العينة الباقية إلى قسمين رئيسيين يمثل القسم الأول من البيانات القيم المنخفضة والقسم الثاني القيم المرتفعة.

D – يقدر نموذج الانحدار لكل قسم على حدة .

E – يحتسب تباين الخطأ العشوائي (مجموع مربعات الخطأ العشوائي aui^2 لكل تقدير ووفقا للصيغة الآتية:

$$ou1^2 = \frac{\sum_{I=}^{T1} UiI^2}{T1-2} \ldots\ldots ou2^2 = \frac{\sum_{I=}^{T2} Ui2^2}{T2-2}$$

$T2$، $T1$: حجم العينة الكلي للقسمين الأول والثاني.

2: عدد المعالم في النموذج

F – يحتسب اختبار] وفقا للصيغة الآتية : $F = \frac{ou\ 2^2}{ou1^2}$

ويتم مقارنة القيمة المحتسبة لاختبار F وفقا للصيغة أعلاه مع القيم الجدولية عند مستوى المعنوية المعين وبدرجة حرية $Ti-2$ فإذا كانت القيمة المحتسبة أقل من القيمة الجدولية فنقبل فرضية العدم HO ويدل ذلك على تباين متجانس والعكس الصحيح .

المحور السادس:

عموميات حول السلاسل الزمنية وتقدير مركباتها

1. ماهية السلاسل الزمنية:

في أدبيات الإحصاء يمكن التعرف على ثلاثة أنواع مختلفة من البيانات، هي: البيانات التجريبية وبيانات الحصر (المسح) والبيانات الزمنية. وتعتمد الفلسفة الخاصة بالبيانات التجريبية على الأسلوب التجريبي والذي يبدأ بتحديد العوامل الهامة والتي يعتقد الباحث أن لها تأثير معنوي على الظاهرة أو المتغير موضع الدراسة، ثم يتم الحصول على البيانات من خلال تصميم تجربة تعتمد على مبدأ العشوائية تسمح بقياس تأثير أحد أو بعض هذه العوامل على الظاهرة أو المتغير موضع الدراسة في ظل ثبات العوامل الأخرى (شعراوي، 2005 صفحة 10). وتعتمد الفلسفة الخاصة ببيانات الحصر أو المسح على مبدأ الحصول على البيانات عن طريق حصر أو مسح الوضع القائم للظواهر موضع الدراسة كما هو دون محاولة التحكم في العوامل المختلفة التي قد تؤدي إلى الحالة التي توجد عليها هذه الظواهر. أما البيانات الزمنية فيتم الحصول عليها من خلال رصد البيانات أو القيم التي تعبر عن الظاهرة أو المتغير موضع الدراسة على فترات زمنية متتالية بهدف تحقيق عدة أهداف أهمها اكتشاف نمط التطور التاريخي للظاهرة أو المتغير موضع الدراسة، وكيفية الاستفادة من هذا النمط في التنبؤ بهذه الظاهرة في المستقبل. ويطلق على البيانات الزمنية "السلاسل الزمنية" وهي الموضوع الرئيسي لهذا الكتاب.

تعريف: السلسلة الزمنية هي مجموعة من المشاهدات أو القياسات التي تؤخذ على إحدى الظواهر الاقتصادية – الاجتماعية – الطبية – الطبيعية – ...... على فترات زمنية متتابعة عادة ما تكون متساوية الطول.

ويمكن رصد السلاسل الزمنية في شتى أنواع المعرفة وميادين التطبيق المختلفة. ففي الاقتصاد يمكن رصد بيانات الدخل السنوي وقيمة التحويلات الخارجية السنوية والإيداعات ربع السنوية في أحد البنوك وعدد العاطلين الشهري وغيرها من البيانات. وفي علم الاجتماع يمكن رصد عدد الجرائم الأسبوعي وعدد حالات الطلاق أو الزواج السنوي وغيرها. وفي مجال التعليم يمكن رصد السلاسل الزمنية الخاصة بتطور أعداد الطلبة السنوي في مراحل التعليم المختلفة وأعداد المدارس والمدرسين السنوية في الكليات المختلفة. وفي مجال الطب يمكن رصد السلاسل الزمنية الخاصة بتطور الأمراض المختلفة ومدى التزايد أو التناقص في الإصابة بهذه الأمراض مثل التطور التاريخي لنسبة المصابين بالذبحة الصدرية أو الأورام الخبيثة، كما يمكن رصد السلاسل الزمنية الخاصة برسم القلب أو الدماغ. وفي مجال الأرصاد الجوية يمكن رصد السلاسل الزمنية الخاصة بكمية الأمطار الشهرية والسلاسل الزمنية الخاصة بسرعة الرياح ونسبة الرطوبة ودرجات الحرارة وغيرها. وفي مجال البيئة يمكن رصد السلاسل الزمنية الخاصة بتطور نسب التلوث في الأجواء المحيطة وتطور متوسط الحموضة في مياه الأمطار السنوية ونسب الأكسوجين المذاب في المياه كمقياس لتلوث المياه. وفي مجال الزراعة يمكن رصد السلاسل

الخاصة بتطور الإنتاج السنوي من المحاصيل الزراعية والدخل السنوي الناتج من قطاع الزراعة. وفي مجال الكيمياء يمكن رصد درجة الحرارة التي تؤخذ كل دقيقة من عملية كيميائية معينة، وفي مجال الهندسة يمكن رصد تطور نسب الوحدات المعيبة الشهرية وتطور إنتاجية العامل السنوية في أحد المصانع.

وتختلف السلسلة الزمنية عن البيانات التجريبية وبيانات الحصر في ثلاث نقاط أساسية هي:

1. تؤخذ بيانات السلسلة الزمنية على فترة زمنية طويلة نسبيًا يعتقد أنها تؤثر على الظاهرة أو المتغير موضع الدراسة، بينما تؤخذ البيانات التجريبية أو بيانات الحصر (المسح) عند نقطة زمنية معينة أو على الأكثر في فترة زمنية قصيرة يعتقد أنها لا تؤثر بشكل معنوي على الظاهرة أو المتغير موضع الدراسة. وعادة ما تسم هذه البيانات بالبيانات المتقاطعة cross sectional data

2. يتم دراسة السلسلة الزمنية عادة بمعزل عن العوامل الأخرى – بخلاف الزمن – التي قد تؤثر عليها وعن الظواهر الأخرى التي قد ترتبط معها في علاقة إحصائية.

3. عادة ما تكون بيانات أو مشاهدات السلسلة الزمنية مرتبطة ببعضها البعض، ويأخذ الارتباط بين هذه المشاهدات أشكالاً وأنماطاً عديدة تختلف باختلاف طبيعية الظاهرة، ومن ثم فإن ترتيب المشاهدات في السلاسل الزمنية ذو أهمية خاصة، ولذلك فإن معظم الأساليب التي تستخدم في تحليل البيانات التجريبية أو بيانات الحصر لا تكون صالحة لتحليل السلاسل الزمنية، وبالتالي كان لابد من ابتكار وتطوير أدوات وأساليب خاصة لتحليل السلاسل الزمنية (شعراوي، 2005 صفحة 9).

وقد خصّص الإحصاء مجالاً منفردًا لتحليل البيانات الزمنية يعرف بمجال لسلاسل الزمنية والذي تطور تطوراً هائلاً في العقود الثلاثة الأخيرة من القرن العشرين بسبب المنهجية الحديثة التي قدمها العالمان بوكس G.E.Box جينكنز Jenkins ... في سنة 1976 والتي يمكن اعتبارها بحق البداية الحقيقية لتحليل الحديث للسلاسل الزمنية والسبب الحقيقي وراء القفزات العلمية الهائلة التي حدثت في هذا المجال. وللمزيد من التفاصيل حول أنواع البيانات الإحصائية والفروق الأساسية بينهم يمكن للقارئ الرجوع إلى شعراوي وإسماعيل (2002).

وتعرض السلسلة الزمنية عادة في صورة جدول أو خط أو منحنى بياني يعرف الخط التاريخي أو المنحنى الزمني time series plot كما في الأمثلة الآتية:

مثال (1): يوضح الجدول الآتي حجم الإنتاج السنوي للبترول بالمليون متر مكعب في إحدى الدول من سنة 1990 إلى سنة 1997:

جدول (1): الخط البياني (التاريخي) لبيانات المثال (1)

السنة	1990	1991	1992	1993	1994	1995	1996	1997
حجم الإنتاج	5210	5820	5655	7800	8100	8010	9200	9335

ويمكن عرض هذه البيانات في شكل خط بياني (تاريخي) في شكل (1):

مثال (2): يوضح الجدول الآتي تطور قيمة الإيداعات ربع السنوية بالمليون دولار في أحد البنوك في سنتي 1998 و 1999:

شكل (2): العرض البياني لبيانات المثال(2)

الإيداعات	الموسم (الفصل)	السنة
42	1	1998
46	2	
50	3	
56	4	
45	1	1999
49	2	
54	3	
60	4	

ويمكن عرض هذه البيانات في شكل (2)

2 أنواع السلاسل الزمنية:

عند دراسة السلاسل الزمنية لبعض الظواهر قد يكون من الممكن أخذ قياسات أو قراءات عند كل لحظة زمنية ويقال لهذه السلاسل بأنها سلاسل متصلة continuous، ومن أمثلة هذه السلاسل درجات الحرارة ورسم القلب ورسم الدماغ. أما معظم السلاسل الزمنية التي تنشأ في الواقع فتتكون من قراءات أو مشاهدات مأخوذة عند فترات زمنية محددة مسبقا. وقد تكون هذه الفترات دقائق أو ساعات أو أيام أو أسابيع أو شهور أو سنوات. وتعرف هذه السلاسل بالسلاسل المتقطعة discrete time series بغض النظر عن طبيعة الظاهرة أو المتغير موضع الدراسة، أمثلة هذه السلاسل الدخل القومي السنوي وسعر الإقفال اليومي لأحد الأسهم فـي بورصة الأوراق المالية وعدد الحوادث الأسبوعية التي تحدث على أحد الطرق وعدد خريجي إحدى الكليات السنوي وكمية الأمطار الشهرية والسلاسل الزمنية المتقطعة السلاسل التي سنتعامل معها فقط في هذا الكتاب، أي أننا سنفترض دائمًا أن السلسلة متاحة فقط عند نقاط زمنية متقطعة تبعد عن بعضها فجوات زمنية متساوية الطول.

وفي الواقع يمكن الحصول على السلسلة الزمنية المتقطعة بمعاينة سلسلة زمنية متصلة وذلك بأن يتم رصد أو تسجيل القراءات فقط عند نقاط زمنية محدَّدة متساوية الأبعاد، كما يمكن الحصول على السلسلة الزمنية

المتقطعة بتراكم متغير معين خلال فترة زمنية مثل كمية الأمطار التي عادة ما تتراكم خلال يوم أو شهر مثلاً أو الناتج السنوي من أحد المحاصيل الزراعية والاهتمام الأساسي لهذا الكتاب هو كيفية بناء النماذج للسلاسل الزمنية المتقطعة واستخدام هذه النماذج في التنبؤ بالمشاهدات المستقبلية.

3.أهداف دراسة السلاسل الزمنية :

تدرس السلاسل الزمنية عادة لتحقيق عدد من الأهداف. وقد يكون أول أهداف هذه الدراسة هو استخدام السلسلة الزمنية لوصف وتصوير المعلومات المتاحة عن فترة زمنية توضح تطور الظاهرة المدروسة أي وصف الملامح والسمات الرئيسية للسلسلة. ويساعد وصف السلسلة إلى حد كبير في تحديد النموذج الذي يمكن أن يكون مناسبًا لتحقيق الأهداف الأخرى والتعرف على حركات الصعود والهبوط في السلسلة الزمنية والتعرف على المكونات الرئيسية مثل الاتجاه العام والتغيرات الموسمية كما سنرى في نهاية هذا الباب. أما الهدف الثاني من دراسة السلاسل الزمنية فهو التفسير، ويقصد به توضيح وشرح التغيرات التي تحدث في الظاهرة باستخدام السلاسل الزمنية الأخرى التي ترتبط بها أو باستخدام عوامل البيئة المحيطة بالظاهرة ومثال ذلك تفسير التغيرات التي تحدث في سلسلة المبيعات الخاصة بإحدى السلع باستخدام السلسلة الخاصة بتغيرات أسعار هذه السلعة أو بمعرفة القرارات الاقتصادية التي اتخذت وكانت لها علاقة مباشرة على التطور التاريخي للظاهرة أو تفسير التغيرات التي تحدث في سلسلة عدد الحوادث التي تحدث على طريق جدة/ مكة المكرمة بالسلسلة الزمنية لعدد الحجاج السنوي أو السلسلة الزمنية لعدد المعتمرين الشهري أو الإجراءات الأمنية التي اتخذت للحد من هذه الحوادث وربط التغيرات التي تحدث في السلسلة موضع الدراسة بالتغيرات التي تحدث في السلاسل والعوامل المحيطة من شأنه فهم آلية عمل السلسلة وتفسير الأنماط والتغيرات المنتظمة وغير المنتظمة التي تتعرض لها الظاهرة موضع الدراسة ومدى تأثير كل منها عليها. والهدف الثالث من دراسة السلاسل الزمنية هو الرقابة والتحكم، فقد تستخدم الخرائط الزمنية في مراقبة جودة الإنتاج وذلك من أجل التحكم في مستوى كفاءة العملية الإنتاجية وذلك باتخاذ القرارات المناسبة من وقف العملية الإنتاجية وتعديل مسارها أو استمرارها (شعراوي، 2005 صفحة 20).

أما أهم أهداف دراسة السلاسل الزمنية على الإطلاق فهو التنبؤ بالمشاهدات المستقبلية والذي عادة ما يمثل الهدف النهائي من تحليل السلاسل الزمنية. وهذا الهدف هو أوضح الأهداف وأكثرها شعبية بالنسبة لدارس الإحصاء أو مستخدمه والذي من أجله كتبت عشرات الكتب العالمية والآلاف من الأبحاث المتخصِّصة. فتحليل السلاسل الزمنية يبدأ عادة بالتعرف على النمط المناسب لشرح آلية تطور هذه السلسلة واستكمال هذا النمط مستقبلاً. والفرض الأساسي في أساليب التنبؤ المستخدمة هو أن هذا النمط الذي تم التعرف عليه سيستمر في

المستقبل القريب، وتجدر الإشارة إلى أنه لا يمكن لأي أسلوب تنبؤ أن يعطي نتائج جيدة إذا لم يستمر هذا النمط، ولذلك فإنه ينصح دائمًا بالتنبؤ بالقيم المستقبلية القريبة وتحديثها بمجرد الحصول على أي مشاهدة جديدة.

4. قياس أخطاء التنبؤ:

عادة ما تدرس السلسلة الزمنية بغرض اكتشاف نمط التطور التاريخي للظاهرة واستغلال هذا النمط في التنبؤ بالقيم المستقبلية. وأي تنبؤ مستقبلي لأي ظاهرة لابد أن يحتوي على قدر معين من عدم التأكد، ويمكن ترجمة هذه الحقيقة بإدراج مركبة خطأ هي error component في نموذج التنبؤ. ومركبة الخطأ هي المركبة غير النمطية التي تعبر عن العوامل التي لا يمكن شرحها باستخدام التغيرات النمطية أو المنتظمة في السلسلة. وكلما كانت هذه المركبة صغيرة زادت قدرتنا على التنبؤ والعكس صحيح (والتر، 1996 صفحة 114). إذا افترضنا أن قيمة الظاهرة موضع الدراسة عند الزمن t هي y_1 وأن التنبؤ بالظاهرة عند الزمن t هو $\hat{y}_1$، فإن الخطأ في التنبؤ عند الزمن t يعرف كالآتي: $e_1 = y_1 - \hat{y}_1 \quad ; \quad t = 1,\ 2, \dots\dots, n$

حيث يرمز n إلى طول السلسلة أي عدد المشاهدات.

وفحص أخطاء التنبؤ المتتالية، يوضح مدى ملاءمة أسلوب التنبؤ المستخدم، فكما هو معروف من دراسة الانحدار أن أسلوب التنبؤ الملائم لابد أن ينتج أخطاءً تتصف بطابع العشوائية، أي أخطاء خالية من أي تغيرات منتظمة –كما في شكل (3) – بالإضافة إلى بعض الشروط الأخرى. وإذا كانت هذه الأخطاء محتملة بحيث يمكن اعتبار أسلوب التنبؤ ملائم فإنه يجب قياس حجم هذه الأخطاء لتقدير دقة التنبؤ.

شكل (3): أخطاء عشوائية

ولقد عرف الفكر الإحصائي طرقا عديدة لقياس حجم الأخطاء أهمها ما يلي:

1– مجموع الأخطاء sum of errors ويرمز له عادة بالرمزSE ويعرف على الصورة الآتية:

$$SE = \sum_{t=1}^{n} e_1 = \sum_{t=1}^{n} (y_1 - \hat{y}_1)$$

وهذا المقياس لا يفيد كثيرًا، حيث أنه من المعروف أنه إذا كانت الأخطاء عشوائية فإن هذا المجموع عادة ما يكون قريبًا جدًا من الصفر بغض النظر عن حجم هذه الأخطاء.

2– متوسط الانحرافات المطلقة mean absolute deviation والذي يرمز له عادة بالرمز MAD ويعرف على الصورة الآتية: $MAD = \frac{1}{n}\sum_{t=1}^{n}|e_1 = \sum_{t=1}^{n} |y_1 - \hat{y}_1|$

وبالرغم من معقولية هذا المقياس إلا أنه لا يستخدم كثيرًا في مجالات السلاسل الزمنية نظرا لصعوبة خصائصه الإحصائية.

3− متوسط مربعات الأخطاء mean squared error ويرمز له عادة بالرمزMSE ويعرف على الصورة الآتية:

$$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2 = \frac{1}{n}\sum_{t=1}^{n}(y_1 - \hat{y}_1)^2$$

ويلاحظ أن طريقة المربعات الصغرى least squares method المعروفة في مجالات الانحدار والسلاسل الزمنية تعتمد على تصغير مجموع مربعات الأخطاء SSE أو تصغير متوسط مربعات الأخطاء MSE وذلك لأن المقام n والذي يمثل عدد الوحدات الزمنية المتاحة (عدد المشاهدات) هو مقدار ثابت وبصفة عامة يمكن القول بأن خصائص هذا المقياس الإحصائية أسهل كثيراً من خصائص متوسط الأخطاء المطلقة MAD

4− متوسط الأخطاء النسبية المطلقة mean absolute percentage error والذي يرمز له عادة بالرمز MAPE ويعرف على الصورة التالية:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_1 - \hat{y}_1}{y_1}\right| = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\mathrm{e}_1}{y_1}\right|$$

ويتميز هذا المقياس عن كل المقاييس بأنه مقياس نسبي، أي لا يعتمد على الوحدات المستخدمة ولكن خصائصه الإحصائية أصعب من خصائص متوسط مربعات الأخطاء MSE ولذلك فإن هذا المقياس عادة ما يستخدم في الدراسات الوصفية التي لا تستدعي القيام باستدلالات إحصائية.

5. اختيار أسلوب التنبؤ المناسب:

من أهم عناصر تحليل السلاسل الزمنية اختيار أسلوب التنبؤ المناسب. واختيار أسلوب التنبؤ المناسب ليس بالعمل الهين وإنما هو عمل صعب وشاق ويحتاج من الإحصائي ومتخذ القرارات التحلي بالصبر وعدم اليأس بالإضافة إلى مقومات العمل الأساسية من علم وخبرة ومهارة. ويعتمد الإحصائي أو متخذ القرارات بصفة عامة في اختياره لأسلوب التنبؤ المناسب على بعض المعايير أو العوامل العامة أهمها ما يلي (والتر، 1996 صفحة 123):

1. تصغير حجم أخطاء التنبؤ أول هذه المعايير التي يجب أن يضعها الإحصائي أو متخذ القرارات نصب عينيه عند اختياره أسلوب التنبؤ، ومن ثم نموذج التنبؤ المناسب، وعادة ما يقاس حجم هذه الأخطاء بأحد المقاييس الثلاثة التي سبق ذكرها (MAD- MSE- MAPE).
2. نوعية التنبؤ المطلوب، فإذا كان تنبؤ النقطة point forecast هو المطلوب من الدراسة، فإن استخدام أحد الأساليب أو النماذج التقليدية البسيطة قد يكون كافيا لتحقيق هذا الهدف. وفي الكثير من الدراسات قد يكون تنبؤ الفترة forecast interval هام وكذلك اختبارات الفروض، وفي مثل هذه الحالات لابد من استخدام أسلوب تنبؤ حديث أكثر دقة وتنظيمًا مثل أسلوب بوكس وجينكنز.

3. عدد المشاهدات المتاحة، فإذا كان عدد المشاهدات صغيرًا فإن استخدام أحد الأساليب الحديثة ليس له ما يبرره ويفضل استخدام أحد الأساليب التقليدية.
4. تكاليف أسلوب التنبؤ ومدى توافر البرامج الإحصائية ذات الصلة.
5. سهولة العمليات الإحصائية والحسابية الضرورية وفهم أسلوب التنبؤ المستخدم.
6. مدى تحقق الفروض النظرية التي يعتمد عليها أسلوب أو نموذج التنبؤ المناسب وهو أهم المعايير التي يجب أن تؤخذ في الاعتبار عند اختيار مثل هذا الأسلوب.

مما سبق يتضح للقارئ بأن أفضل أسلوب للتنبؤ ليس بالضرورة هو الأسلوب الذي يحقق أعلى دقة أو أصغر حجم أخطاء ممكن، فقد يستخدم أحد الأساليب بسبب نوعية التنبؤ المطلوب، وقد يستخدم أسلوب آخر بسبب صغر عدد المشاهدات المتاحة، وقد يستخدم أسلوبًا ثالثًا بسبب انخفاض تكاليفه، وقد يستخدم أسلوب رابع بسبب سهولة عملياته الإحصائية والحسابية، وقد يستخدم أسلوبًا خامسًا لأن الفروض النظرية التي يعتمد عليها تتوافق مع بيانات السلسلة المتاحة. وعادة ما يعتمد أسلوب التنبؤ المستخدم على قدرة الإحصائي أو متخذ القرارات في تحقيق التوازن لكل هذه المعايير. وبصفة عامة يمكن القول بأن طريقة التنبؤ التي يجب استخدامها هي أسهل وأبسط طريقة يمكن تنفيذها في الزمن المتاح والتي تفي باحتياجات وظروف التنبؤ بأقل تكاليف ممكنة، وللمزيد من التفاصيل حول قياس الأخطاء ومعايير التنبؤ يمكن للقارئ الرجوع إلى Bowerman and O'Connell (1987) أو إلى Gaynor Kirkpatrick and (1994)

6. طرق التنبؤ:

يمكن تجميع طرق التنبؤ الكمية المعروفة في أدبيات السلاسل الزمنية في أسلوبين أساسيين هما.

أ- أسلوب الانحدار regression approach, ويعتمد هذا الأسلوب على تحديد المتغيرات الأخرى التي قد ترتبط بعلاقة سببية بالظاهرة أو المتغير موضع الدراسة الذي يراد التنبؤ به والذي عادة ما يعرف بالمتغير التابع dependent variable – ثم تحديد النموذج الإحصائي أو العلاقة الدالية الملائمة التي توضح الكيفية التي يرتبط بها هذا المتغير بالمتغيرات الأخرى والتي تأخذ في العرف الإحصائي أسماء عديدة مثل المتغيرات المستقلة independent variables أو المتغيرات المفسرة regressors أو المتغيرات المنبئة predictors. وباستخدام هذا النموذج يمكن التنبؤ بالمتغير التابع موضع الدراسة في المستقبل إذا أمكن تحديد أو معرفة القيم المستقبلية للمتغيرات المفسرة. وتعرف النماذج التي تندرج تحت مظلة هذا الأسلوب أحيانا بالنماذج السببية causal models ويستخدم هذا الأسلوب في كافة أنواع المعرفة ومجالات التطبيق الخاصة الاقتصادية والاجتماعية والبيئية منها حيث يسمح هذا الأسلوب بتقييم أثر المتغيرات المتضمنة والتي عادة ما تعكس أثر الأنظمة والسياسات والقرارات

المختلفة، فعلى سبيل المثال قد نستطيع تفسير السلسلة الخاصة بقيمة المبيعات اليومية لإحدى السلع جزئيًا بواسطة سلسلة أسعار هذه السلعة والسلاسل الخاصة بالدخل الفردي وأسعار السلع البديلة. وهذا الأسلوب بالرغم من شعبيته يعاني من بعض العيوب أهما ما يلي:

1– صعوبة تحديد المتغيرات المفسرة التي ترتبط بالمتغير التابع أو الظاهرة موضع الدراسة.

2– تطبيق هذا الأسلوب يتطلب توافر بيانات تاريخية تفصيلية عن جميع المتغيرات المفسرة والقدرة على معرفة قيم هذه المتغيرات – أو على الأقل التنبؤ بها– عند الأزمنة التي يراد التنبؤ بالظاهرة عندها.

3– يفترض عدم الارتباط بين مشاهدات المتغير أو الظاهرة موضع التنبؤ، وهو فرض غير واقعي ولا يتفق مع مفهوم السلسلة الزمنية باعتبارها مجموعة من المشاهدات المرتبطة وعادة ما يؤدي هذا الفرض غير الواقعي إلى تنبؤات غير موثوق بها.

وعادة ما يتوافر لدى الباحث مشاهدات تاريخية عن المتغير موضع الدراسة فقط ويريد التنبؤ بالمشاهدات المستقبلية لهذا المتغير بالاعتماد فقط على هذه المشاهدات. في مثل هذه الحالات يستخدم الأسلوب الثاني للتنبؤ التالي.

ب – تحليل السلاسل الزمنية time series analysis والذي يضم تحت مظلته ما يعرف بنماذج السلاسل الزمنية، ويعتمد هذا الأسلوب على تحليل البيانات التاريخية التي أخذت عن الظاهرة أو المتغير موضع الدراسة وذلك بغرض تحديد نمط البيانات. بعد ذلك – وبافتراض أن هذا النمط سيستمر في المستقبل – يستكمل هذا النمط لإعطاء التنبؤات المطلوبة. فعلى سبيل المثال إذا كان الهدف من الدراسة التنبؤ بعدد الحجاج السنوي فقد يستطيع الباحث تفسير سلوك هذا المتغير جزئيا بواسطة عدد السكان في الدول الإسلامية ومتوسط دخل الفرد في هذه البلاد ومتوسط التكاليف المطلوبة لأداء هذه الفريضة، ولكن جزء كبير من تطور عدد الحجاج قد يعود إلى بعض العوامل الأخرى التي لا يمكن أخذها في الاعتبار بسهولة مثل الوازع الديني وحالة الطقس وغيرها من العوامل التي يكون من الصعب أو المستحيل إدراجها في النماذج السببية بسهولة. وفي هذه الحالة قد يفضل دراسة التطور التاريخي لعدد الحجاج السنوي بمعزل عن جميع العوامل المفسرة الأخرى واكتشاف الكيفية التي يتطور بها عدد الحجاج واستخدام أحد نماذج السلاسل الزمنية لاستكمال هذه السلسلة في المستقبل.

وأسلوب السلاسل الزمنية والذي يضم ما يعرف بنماذج أو طرق السلاسل الزمنية هو محور الاهتمام الرئيسي لهذا الكتاب، ولن نتعرض للنماذج السببية بأي حال من الأحوال. والسؤال الآن هو كيف يمكن التنبؤ بالظاهرة أو المتغير موضع الدراسة باستخدام أسلوب السلاسل الزمنية دون اللجوء إلى أي متغيرات أخرى مفسرة؟ للإجابة عن

هذا السؤال يمكن القول بأن أدبيات السلاسل الزمنية قد عرفت العديد من الطرق ونماذج السلاسل الزمنية والتي يمكن تقسيمها إلى ثلاثة أنواع رئيسية هي:

- النماذج المحددة (غير العشوائية) deterministic models
- والطرق الحسية ad hoc methods
- ونماذج السلاسل الزمنية العشوائية stochastic time series models

ونقدم فيما يلي عرضًا سريعًا لهذه الطرق والنماذج.

1.6 النماذج المحددة Deterministic Models

نعرف من دراستنا في علم الإحصاء أن نموذج المتوسط mean model يمكن التعبير عنه في الصورة العامة

$$y_1 = E(y_1) + \varepsilon_1$$

حيث ε_1 متغيرات عشوائية غير مرتبطة توقعها الصفر وتباينها ثابت ويقال أن هذا النموذج محدد deterministic أو غير عشوائي nonstochastic إذا أمكن التعبير عن$E(y_1)$ كدالة رياضية مباشرة في الزمن t ولتكن $f(t,\beta)$ حيث يرمز المتجه β إلى معالم هذه الدالة الرياضية. وفي هذه الحالة يمكن التعبير عن مشاهدات السلسلة الزمنية y_1 على الصورة:

$$y_1 = f(t,\beta) + \varepsilon_1 \quad ; \quad t = 1,2,\dots,n \qquad (1.6.2)$$

ويعني هذا أن المشاهدات المستقبلية للسلسلة يمكن التعبير عنها على الصورة:

$$y_h = f(h,\beta) \quad ; \quad h = t+1\,, \qquad t+2, \dots\dots$$

أي أن هذه النماذج تفترض أن المشاهدات المستقبلية للسلسلة تأخذ شكلاً رياضيًا محددًا، أي غير عشوائي

وتعتمد النماذج المحددة (1.6.2) على فرضيين أساسيين. الفرض الأول أن الدالة $f(t,\beta)$ دالة رياضية محددة ليس لها طابع العشوائية، والفرض الثاني أن ε_1 متغيرات عشوائية غير مرتبطة توقعها الصفر وتباينها ثابت وتؤدي هذه الفروض إلى أن المتغيرات $y_1\,, y_2\,, \dots, y_n$ تكون متغيرات عشوائية غير مرتبطة، ومن أمثلة الدوال الرياضية التي تستخدم في هذه النماذج كثيرات الحدود والدوال الأسية والدوال المثلثية، وفيما يلي عرضا مبسطا لكثيرات الحدود والدوال الأسية.

1.1.6 كثيرات الحدود Polynomials:

ويفترض هنا أن الدالة f(t) أي متوسط الظاهرة تأخذ إحدى صور كثيرات الحدود في الزمن t وتعتبر الصورة الخطية أهم هذه الصور وتعرف على الشكل الآتي:

$$E(y)_1 = f(t) = \beta_0 + \beta_1 t$$

وتكون هذه الدالة ملائمة إذا أمكن تمثيل متوسط الظاهرة بواسطة خط مستقيم وذلك بعد توقيع مشاهدات السلسلة على ورقة الرسم البياني كما في شكل (4) .

وتفترض كثيرة الحدود الخطية أن متوسط الظاهرة يتزايد بمعدل ثابت β_1 كما في الشكل (4.a) أو يتناقص بمعدل ثابت β_1 كما في الشكل (4.b)

ويمكن إيجاد تقديري المربعات الصغرى $\hat{\beta}_1, \hat{\beta}_2$ لمعلمتي النموذج β_0, β_1 بإجراء انحدار السلسلة y_1 على الزمن t، ومن ثم يمكن التنبؤ بالمشاهدات المستقبلية باستخدام النموذج المقدر الآتي: $\hat{y}_1 = \hat{\beta}_1 + \hat{\beta}_2$

وفي بعض الأحيان قد يكون الشكل الخطي غير ملائم لتمثيل متوسط الظاهرة أو الدالة $f(t)$ ومن الأفضل تمثيل هذا المتوسط بكثيرة حدود من الدرجة الثانية كما في الشكل (5)– والتي تأخذ الصورة الآتية:

$$E(y_1) = f(t) = \beta_0 + \beta_1 t + \beta_2 t^2$$

وتستخدم طريقة المربعات الصغرى العادية لإيجاد التقديرات $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ وذلك بإجراء الانحدار الخطي المتعدِّد لقيم السلسلة y_1 على المتغيرين t, t^2. وسنوضح هذا بالتفصيل عند قياس الاتجاه العام باستخدام تحليل الانحدار.

2.1.6 النمو الأسي Exponential Growth

في بعض الأحيان قد يفضل تمثيل متوسط الظاهرة موضع الدراسة في شكل دالة أسية – كما في الشكل (6)–على الصورة:

$$E(y_1) = f(t) = ce^{rt} \qquad (1.6.3)$$

حيث c و r مقداران ثابتان يمثلان معلمتي النموذج.

ويفترض النموذج الأسى أن متوسط الظاهرة ينمو بنسبة ثابتة وذلك لأن من (1.6.3):

$$\frac{E(y_1)}{E(y_{t-1})} = \frac{ce^{rt}}{ce^{r(t-1)}} = e^r$$

حيث e^r ثابت لا يعتمد على الزمن، ويعني هذا أن: $E(y_1) = e^r E(y_{t-1})$

ويمكن تحويل الصورة (1.6.3) إلى الصورة الخطية بأخذ اللوغاريتمات الطبيعية للطرفين كما يلي:

$\ln f(t) = \ln c + rt$ أي أن $f^*(t) = c^* + rt$ حيث: $f^*(t) = \ln f(t)$; $c^* = \ln c$

وبالتالي يمكن إجراء انحدار لوغاريتم البيانات الأصلية على الزمن t وإيجاد تقديري المربعات الصغرى للمعلمتين c^*, r ومن ثم يمكن إيجاد تقدير الثابت c من العلاقة: $c = e^{c^*}$

ومن ثم يمكن التنبؤ بالمشاهدات المستقبلية باستخدام النموذج المقدر الآتي: $\hat{y}_1 = \hat{c}\, e^{\hat{r}t}$

وتعاني الطرق أو النماذج المحددة في تحليل السلاسل الزمنية من العديد من العيوب أهمها:

1– تركز هذه الطرق على المنطق الرياضي في محاولة لإيجاد دالة رياضية جيدة ممكن أن تستخدم في توفيق البيانات أكثر من اهتمامها بمحاولة استكشاف الخصائص الإحصائية الهامة للسلسلة وأهمها نمط الارتباط الموجود بين مشاهدات السلسلة، فهذه النماذج لا تصف خصائص السلسلة الإحصائية ولكنها مجرد نماذج تنتج المشاهدات $y_1, y_2, \dots. y_n$

2– تفترض هذه الطرق أن التطور طويل الأجل في السلسلة يكون نمطي أو منتظم الشكل وممكن التنبؤ به بشكل كبير.

3– وتفترض أيضًا هذه الطرق عدم وجود ارتباط بين مشاهدات السلسلة، وهذا الفرض من النادر أن يكون متحققا في مجالات التطبيق المختلفة. بسبب كل هذه العيوب فإن هذه الطرق عادة ما تؤدي إلى تنبؤات غير دقيقة من الناحية الإحصائية.

وبالرغم من الانتقادات التي توجه إلى هذه الطرق فإن لها باع طويل في موضوعات التنبؤ ومازالت تستخدم بكثرة في كافة مجالات التطبيق في البلاد النامية خاصة في الاقتصاد والإدارة والبيئة لأنها وسائل بسيطة وغير مكلفة ولا تحتاج إلى خبرات أو مهارات خاصة من قبل الباحثين أو الدارسين أو متخذي القرارات.

2.6 الطرق الحسية Ad Hoc Methods:

ذكرنا أن طرق التنبؤ المحددة تعتمد على التعبير عن قيمة السلسلة عند الزمن t أي y_1 كدالة رياضية مباشرة في الزمن. وهذا الاتجاه له عيوبه كما ذكرنا وأهمها أنه يفترض عدم وجود علاقة بين مشاهدات السلسلة $y_1, y_2, \dots. y_n$, الاتجاه الثاني في التنبؤ يعتمد على التعبير عن تنبؤ السلسلة عند الزمن t بدلالة حاضر السلسلة y_1 وماضي السلسلة $y_1, y_2, \dots. y_{t-1}$,. فإذا افترضنا أن t تمثل نقطة أصل معينة وأننا نريد التنبؤ بقيمة السلسلة بعد k من الفترات الزمنية فإن الاتجاه الثاني يفترض العلاقة الدالية الآتية:

$$\hat{y}_{l+k} = f(y_1, y_2, \dots., y_{t-1}, y_t) \qquad (1.6.4)$$

ويوجد العديد من الطرق والنماذج التي تنتمي بشكل أو بآخر للصورة (1.6.4) وتعتمد على الحس الإنساني أكثر من اعتمادها على أسلوب إحصائي منظم. ومن أمثلة ذلك طريقة التنبؤ السطحي وتنبؤ التغير الثابت وطريقة المتوسطات المتحركة البسيطة وطريقة التمهيد الأسي، وفيما يلي نقدم عرضًا مبسطا لهذه الطرق (شعراوي، 2005 صفحة 25).

1.2.6 التنبؤ السطحي Naive Forecasting:

تستخدم طريقة التنبؤ السطحي قيمة المشاهدة الحالية كتنبؤ مباشر للمشاهدة التالية، أي أن:

$$\hat{y}_{t+1} = y_1 \qquad (1.6.5)$$

والنموذج (1.6.5) يكون ملائمًا عندما تكون قيم السلسلة ثابتة بشكل تقريبي على الفترة الزمنية موضع الدراسة، ويحدث هذا عندما يغلب على السلسلة الزمنية محل الدراسة الطابع غير النمطي (غير المنتظم) أي عندما تتغير السلسلة بشكل عشوائي كبير لا يتبع نمطا أو نظامًا أو اتجاها معينًا يمكن معه التنبؤ بقيمة السلسلة في الفترة الزمنية التالية، وتعتبر أسعار الأوراق المالية في البورصة أشهر مثال على هذا النوع من السلاسل الزمنية.

والجدير بالذكر أن النموذج السطحي يعطي تنبؤات متحيزة إلى أسفل إذا كانت السلسلة تتزايد باستمرار وذلك لأن التنبؤ $\hat{y}_1$ يكون دائما أقل من القيمة، y_1 والعكس صحيح، أي أن هذا النموذج يعطي تنبؤات متحيزة إلى أعلى إذا كانت السلسلة تتناقص باستمرار لأن التنبؤ $\hat{y}_1$ يكون أكبر من القيمة y_1 ، ولذلك لا ينصح باستخدام طريقة التنبؤ السطحي في مثل هذه الحالات.

2.2.6 تنبؤ التغير الثابت Constant Change Forecasting:

في الكثير من التطبيقات خاصة الاقتصادية منها تتميز بعض السلاسل بثبات في التغيرات المتتالية، فإذا افترضنا أن t تمثل نقطة أصل معينة فإن التغير السابق في السلسلة يكون: $\Delta y_t = y_t - y_{t-1}$

وإذا كانت $\hat{y}_{t+l}$ تمثل التنبؤ بقيمة السلسلة الزمنية، فإن التغير القادم في السلسلة يقدر كالتالي:

$$\hat{\Delta} y_{1+1} = \hat{y}_{t+1} - y_1$$

وبمساواة التغير السابق Δy_1، بالتغير اللاحق $\hat{\Delta} y_{1+1}$ نصل إلى: $\hat{y}_{t+1} - y_t = y_t - y_{t-1}$

ومن ثم نصل إلى النموذج الآتي:

$$\hat{y}_{t+1} = y_t + (y_t - y_{t-1}) \qquad (1.6.6)$$

أي أن التنبؤ في الفترة الزمنية القادمة يساوي القيمة الحاضرة y_1التغير الذي مضافا إليه قيمة التغير الذي حدث في الفترة السابقة Δy_t.

3.5.6 المتوسطات المتحركة البسيطة Simple Moving Averages: (خالد زهدي صفحة 8)

يعتمد النموذج السطحي على القيمة الحالية y_t فقط للتنبؤ بالقيمة التالية y_{t+1} ، بينما يعتمد تنبؤ التغير الثابت على أحدث قيمتين y_{t-1}, y_t للتنبؤ بالقيمة التالية y_{t-1}... أما طريقة المتوسطات المتحركة البسيطة فتستخدم أحدث k قيمة للسلسلة للتنبؤ بالقيمة التالية أي تستخدم القيم $y_t, y_{t-1}, y_{t-2}, \dots, y_{t-(k-2)}$، وذلك بأخذ متوسط هذه القيم كما يلي:

$$\hat{y}_{t+1} = \frac{1}{k}\left[y_t + y_{t-1} + \cdots + y_{t-(k-2)} + y_{t-(h-1)}\right] \; ; \; t = k, k+1, \dots . n \quad (1.6.7)$$

وهذا يعني أن: $\hat{y}_{t+2} = \frac{1}{k}\left[y_{t+1} + y_1 + y_{t-1} + \cdots + y_{t-(k-2)}\right]$

أي أنـه لإيجـاد المتوسـط المتحـرك البسـيط $\hat{y}_{t+2}$ نستخدم نفس القيم التـي استخدمت فـي حسـاب المتوسـط السابق له مباشرة $\hat{y}_{t+1}$ بعد إحلال القيمة الأحدث y_{t+1} مكان القيمة الأقدم $y_{t-(h-1)}$ وهذا معنى التحرك أي أن المتوسط يتم تحديثه دائمًا بحذف المشاهدة الأقدم ووضـع بـدلاً منها المشاهدة التالية. فعلى سبيل المثال إذا كانت $k = 3$ فإنه يمكن تكوين $(n-3)$ متوسط متحرك بسيط مناظر لقيم السلسلة المتاحة كما يلي:

$$\hat{y}_4 = \frac{1}{3}[y_3 + y_2 + y_1]$$
$$\hat{y}_5 = \frac{1}{3}[y_4 + y_3 + y_2]$$
$$\hat{y}_6 = \frac{1}{3}[y_5 + y_4 + y_3]$$
$$\vdots \qquad \qquad \vdots$$
$$\hat{y}_n = \frac{1}{3}[y_{n-1} + y_{n-2} + y_{n-3}]$$

واختيار العدد الصحيح k يعتمد على رأي الباحث وخبرته العملية وهو أحد المشاكل التي تواجه مستخدم طريقة المتوسطات المتحركة البسيطة. ودقة التنبؤ تعتمد على اختيار العدد الملائم ولذلك يمكن اختيار هذا العدد بطريقة التجربة والخطأ، حيث تحسب جميع التنبؤات التي تناظر كل قيمة من قيم k الممكنة $(k = 2, 3, \dots, n-1)$ وحساب الأخطاء ومن ثم حساب أحد المعايير الهامة لقياس حجم الأخطاء – وليكن متوسط مربعات –الأخطاء المناظر لكل قيمة من قيم k واختيار قيمة k التي تناظر أصغر قيمة لهذا المعيار. وبالرغم من المشاكل التي قد تواجه الباحث عند اختيار قيمة k الملائمة إلا أن العيب الرئيسي لطريقة المتوسطات المتحركة البسيطة هو إعطاء أوزان متساوية لكل المشاهدات المستخدمة في حساب المتوسط فإذا كانت $k = 8$ مثلاً فإن الوزن الذي يعطى للقيمة الحديثة y_t يساوي الوزن الذي يعطى للقيمة الأقدم y_{t-7}. وهذا عادة ما يتعارض مع خصائص السلاسل الزمنية حيث نميل إلى إعطاء المشاهدات الأحدث أوزانا أكبر . ولذلك يفضل استخدام طريقة المتوسطات المتحركة البسيطة عندما يغلب الطابع العشوائي على بيانات السلسلة.

مثال (3): الجدول الآتي يوضح قيمة المبيعات السنوية من إحدى السلع بملايين الدولارات في الفترة من السنة 1990 إلى سنة 1998:

السنة	1990	1991	1992	1993	1994	1995	1996	1997	1998
قيمة المبيعات	9	11	10	12	11	9	13	11	9

a. استخدام طريقة المتوسطات المتحركة البسيطة في إيجاد جميع التنبؤات الممكنة مرة باستخدام $k = 2$ ومرة باستخدام $k = 3$ وأوجد متوسط مربعات الأخطاء المناظرة في كل حالة.

b. تنبأ بقيمة المبيعات في سنة 1999.

الحل:

		K			
		2		3	
السنة	y_t	$\hat{y}_t$	e_1^2	$\hat{y}_t$	e_1^2
1990	9	−	−	−	−
1991	11	−	−	−	−
1992	10	$\frac{1}{2}[9+11]=10$	0	−	−
1993	12	$\frac{1}{2}[10+11]=10.5$	25.2	$\frac{1}{3}[10+1+9]=10$	4
1994	11	$\frac{1}{2}[12+10]=11$	0	$\frac{1}{3}[12+10+11]=11$	0
1995	9	$\frac{1}{2}[11+12]=11.5$	6.25	$\frac{1}{3}[11+12+10]=11$	4
1996	13	$\frac{1}{2}[9+11]=10$	9	$\frac{1}{3}[9+11+12]=10.67$	5.429
1997	11	$\frac{1}{2}[13+9]=11$	0	$\frac{1}{3}[13+9+11]=11$	0
1998	9	$\frac{1}{2}[11+13]=12$	9	$\frac{1}{3}[11+13+9]=11$	4
1199	؟	$0\frac{1}{2}[9+11]=12$	−	$\frac{1}{3}[9+13+11]=11$	−

a. إذا كانت $k=2$: $MSE=\frac{2}{7}[0+2.25+0+6.25+9+0+9]=3.786$

إذا كانت $k=3$: $MSE=\frac{1}{3}[4+0+4+5+5.429+0+4]=2.905$

b. ويمكن القول بأن قيمة k = 3 أفضل من قيمة k = 2 في التنبؤ لأن متوسط مربعات الأخطاء المناظر أقل ومن ثم فإن: $\hat{y}_{10}=\frac{1}{3}[y_9+y_8+y_7]=\frac{1}{3}[9+11+13]=11$

المحور السابع:

الاستقرارية والارتباط الذاتي والجزئي

1. دالة الارتباط الذاتي Autocorrelation Function:

تقيس دالة التغاير الذاتي $\gamma(s,t)$ التي سبق تعريفها في المبحث السابق- درجة الاعتماد الخطي بين أي متغيرين من المتغيرات التي تقع على نفس السلسلة الزمنية. فعلي سبيل المثال يقيس التغاير الذاتي $\gamma(1,2)$ درجة الاعتماد الخطي بين المتغير العشوائي y_1 والذي يمثل قيمة السلسلة عند النقطة الزمنية الأولى والمتغير العشوائي y_2 والذي يمثل قيمة السلسلة عند النقطة الزمنية الثانية، أي أن $\gamma(1,2)$ يمثل درجة الاعتماد الخطي بين كل القيم التي يمكن أن تولدها العملية العشوائية عند النقطة الزمنية الأولى وتلك القيم التي يمكن أن تولدها نفس العملية العشوائية عند النقطة الزمنية الثانية. وبصفة عامة فإن التغاير الذاتي $\gamma(s,t)$ هو دالة في الدليلين s,t .

وتجدر الإشارة هنا إلى بعض الملاحظات الهامة والجديرة بالذكر أهمها:

1- إذا كان $\gamma(s,t)=0$ فهذا يعني أن المتغيرين Y_s,Y_t غير مرتبطين خطيا ولكن قد يكون هناك ارتباط غير خطي بينهما.

2- إذا كان $\gamma(s,t)=0$ وكان المتغيران Y_s,Y_t لهما توزيع معتاد ثنائي Bivariate normal distribution فإن هذا يعني أن المتغيرين مستقلان.

3- يمكن اعتبار تباين العينة كحالة خاصة من دالة التغاير $\gamma(s,t)$ بوضع $s=t$، وهذا يعني أن $V(Y_t)=\gamma(t,t)$

4- إذا كانت السلسلة ساكنة فإن دالة التغاير $\gamma(s,t)$ تكون دالة في الفجوة الزمنية $k=|s-t|$ فقط وتكتب عادة في هذه الحالة $\gamma(|s-t|)$ أو $\gamma(k)$

1.1 ماهية الارتباط الذاتي:

من المعروف في علم الإحصاء أن استخدام التغاير لقياس درجة الاعتماد الخطي بين متغيرين يثير بعض المشاكل العملية، أولها عدم وجود حدود مرجعية (دنيا عليا) يمكن الرجوع إليها لتحديد مدى قوة أو ضعف العلاقة الخطية، وثانيها أن التغاير يعتمد على وحدات القياس المستخدمة (شعراوي، 2005 صفحة 111). ومن ثم يفضل معايرة التغاير الذاتي بالقسمة على حاصل ضرب الانحرافين المعياريين للمتغيرين Y_s,Y_t لنحصل على ما يعرف بالارتباط الذاتي (التسلسلي) .

تعريف:

يعرف معامل الارتباط الذاتي $\rho(s,t)$ بأنه معامل الارتباط الخطي بين المتغيرين Y_s, Y_t ويكتب على الصورة: $\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{Var(Y_s)\,.\,Var(Y_t)}} = \frac{E(Y_s-\mu_s)(Y_t-\mu_t)}{\sqrt{E(Y_s-\mu_s)^2\,.E(Y_t-\mu_t)^2}}\;; s,t = 0 \pm 1 \pm 2, \dots$

وبالطبع – كما هو معروف في علم الإحصاء – يمكن حساب بسط معامل الارتباط من دالة الاحتمال الثنائية للمتغيرين Y_s, Y_t ، بينما يحسب المقام من دالتي الاحتمال الهامشي للمتغيرين وذلك لكل قيم (s,t) المختلفة. وبالتالي ينشأ لدينا علاقة دالية بين معاملات الارتباط الذاتي والزمنين (s,t) تسمى بدالة الارتباط الذاتي(act autocorrelation function) تقيس درجة الارتباط الخطي بين المتغيرات التي تقع على نفس السلسلة أو العملية العشوائية. وتتصف هذه الدالة بعدة خصائص أهمها:

1- الارتباط الذاتي بين المتغير Y_t ونفسه يساوي الواحد، أي أن: $\rho(t,t) = 1$

2- $\rho(t,s) = \rho(s,t)$ وذلك لأنّ $\gamma(t,s) = \gamma(s,t)$

3- قيمة $\rho(s,t)$ تقع دائماً على الفترة المغلقة $[-1,1]$

4- إذا كان $\rho(s,t) = 0$ فهذا يعني أنه لا توجد علاقة خطية بين المتغيرين Y_s, Y_t. ولكن قد توجد علاقة غير خطية بينهما.

5- إذا كان $\rho(s,t) \pm 1$ فهذا يعني أنه يوجد علاقة خطية تامة (طردية أو عكسية) بين المتغيرين Y_s, Y_t ، أي أنه يمكن التنبؤ بأحد المتغيرين بالضبط باستخدام علاقة خطية يقوم فيها المتغير الآخر بدور المتغير المفسر regressor الوحيد في هذه العلاقة.

أما إذا كانت العملية العشوائية (السلسلة) ساكنة فإنه يمكن إعادة تعريف معامل الارتباط الذاتي ودالة الارتباط الذاتي كما يلي:

تعريف: يعرف معامل الارتباط الذاتي للعملية الساكنة $\{y_t\}$ عند الفجوة الزمنية k بأنه معامل الارتباط الخطي بين المتغيرين y_t, y_{t-k} ويأخذ الصورة الآتية:

$$\rho(k) = \frac{E(Y_t-\mu)(Y_{t-k}-\mu)}{E(Y_t-\mu)^2} = \frac{\gamma(k)}{\gamma(0)} \;\;, k = 0, \pm 1 \pm 2 \dots$$

حيث يرمز $\gamma(0)$ إلى تباين العملية الساكنة ويرمز $\gamma(k)$ إلى التغاير الذاتي عند الفجوة k لنفس العملية.

ومن ثم يمكن حساب معامل الارتباط الذاتي لكل فجوة من الفجوات الزمنية $k = 0, \pm 1 \pm 2 \dots$ فينشأ لدينا علاقة دالية بين معاملات الارتباط الذاتي $\rho(k)$ والفجوة الزمنية k تسمى بدالة الارتباط الذاتي للعملية الساكنة $\{y_1\}$ تقيس الارتباط الخطي بين المتغيرات على نفس السلسلة الزمنية والتي تبعد عن بعضها البعض فجوة زمنية مقدارها k. فعلى سبيل المثال يقيس $\rho(1)$ درجة الارتباط الخطي بين أي متغيرين الفاصل الزمني

بينهما يساوي الوحدة أي درجة الارتباط بين Y_1, Y_2، أو بين Y_{11}, Y_{10} أو.. أو بصفة عامة الارتباط الخطي بين Y_t, Y_{t-1}، ويقيس $\rho(3)$ درجة الارتباط الخطي بين أي متغيرين الفاصل الزمني بينهما يساوي ثلاث وحدات أي درجة الارتباط الخطي بين Y_4, Y_1 أو بين Y_{14}, Y_{10} أو...أو بصفة عامة درجة الارتباط الخطي بين Y_t, Y_{t-3} وتعرض دالة الارتباط الذاتي في شكل رياضي أو جدولي أو بياني كما سنرى فيما بعد.

تعريف: تعرف مصفوفة الارتباط الذاتي للعملية العشوائية الساكنة التي تتكون من عدد n من المتغيرات في الصورة الآتية:

$$\begin{bmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(n-1) \\ \rho(1) & 1 & \rho(1) & \cdots & \rho(n-2) \\ \rho(2) & \rho(1) & 1 & \cdots & \rho(n-3) \\ \vdots & \vdots & \vdots & & \vdots \\ \rho(n-1) & \rho(n-2) & \rho(n-3) & \cdots & 1 \end{bmatrix}$$

2.1. خصائص وأهمية دالة الارتباط الذاتي:

تتصف دالة الارتباط الذاتي $\rho(k)$ لأي عملية ساكنة بعدة خصائص هامة نذكر منها ما يلي:

1- معامل الارتباط الذاتي عند الفجوة الزمنية صفر يساوي الواحد، أي أن $\rho(0) = 1$ لأي عملية ساكنة.

2- قيمة $\rho(k)$ تقع دائماً على الفترة المغلقة $[-1,1]$

3- إذا كان $\rho(k) = 0$ فهذا يعني أنه لا توجد علاقة خطية بين أي متغيرين الفاصل الزمني بينهما k وحدة ، ولكن بالطبع قد توجد علاقة غير خطية بينهما.

4- إذا كان $\rho(k) = \pm 1$ فهذا يعني أنه توجد علاقة خطية تامة بين أي متغيرين الفاصل الزمني بينهما k وحدة، أي أنه يمكن التنبؤ بأحدهما بالضبط باستخدام علاقة خطية يقوم فيها المتغير الآخر بدور المتغير المفسر الوحيد في هذه العلاقة.

5- الدالة $\rho(k)$ دالة متماثلة دائماً حول الفجوة $k = 0$، أي أن $\rho(k) = \rho(-k)$. ولذلك عادة ما يكتفي برسم هذه الدالة لقيم k الموجبة فقط كما سنرى فيما بعد في الأمثلة.

6- مصفوفة الارتباط دائما موجبة تامة Positive definite ، ولذلك ترتبط معاملات الارتباط المختلفة $\rho(0), \rho(1), \rho(2)$ .... فيما بينهما بعلاقات جبرية يمكن استنتاجها من العلاقة بين المصفوفة تامة الإيجاب والمحددات الرئيسية كما هو الحال في مصفوفة التغاير والتباين في نظرية الإحصاء.

وتأخذ الدالة $\rho(k)$ في التحليل الحديث للسلاسل الزمنية أشكالاً متعدِّدة، فتارة تجدها تتلاشى ببطء، وتارة تجدها تتلاشى بسرعة في صورة أسية exponential fashion ، وتارة ثالثة تقترب تدريجيا من الصفر في شكل موجات تحاكي دالة الجيب sinewaves، وتارة رابعة تقترب تدريجيًا من الصفر في شكل توليفة من الدوال الأسية، وتارة خامسة تتقطع كلية فجأة بعد عدد معين من الفجوات الزمنية. ففي شكل $(2.a)$ تتناقص $\rho(k)$ ببطء، وتتناقص برتابة وبسرعة وفي صورة أسية في شكل $(2.b)$، وتقترب تدريجياً من الصفر في شكل موجات تحاكي دالة الجيب $(2.c)$، بينما تتقطع كلية فجأة بعد الفجوة الزمنية الثانية في شكل $(2.b)$

وتلعب دالة الارتباط الذاتي $\rho(k)$ دورا هاما وخطيرا – إن لم يكن أهم دور على الإطلاق في التحليل الحديث للسلاسل الزمنية بطريقة بوكس وجينكنز. فهي الأداة الرئيسية التي ارتضاها هذان العالمان لاختبار سكون السلسلة – بجانب الطرق التقريبية الأخرى وهي أحد الأدوات الرئيسية للتعرف على النموذج المبدئي الملائم للسلسلة. بالإضافة إلى ذلك فإن هذه الدالة من أهم أدوات تشخيص النموذج المبدئي من أجل تحسينه أو تطويره إذا ما طبقت على البواقي Residuals الناتجة من هذا النموذج. وسنتعرض للدور الذي تلعبه دالة الارتباط الذاتي بالتفصيل عند تقديم منهجية بوكس وجينكنز في الباب الرابع (شعراوي، 2005 صفحة 120).

شكل (2): بعض دوال الارتباط الذاتي المشهورة

والأمثلة الآتية توضح كيفية إيجاد الارتباط الذاتي لبعض العمليات العشوائية والنماذج ذات الاتجاه المحدد (غير العشوائي).

<u>**مثال (6):**</u> أوجد دالة الارتباط الذاتي لعملية "الاضطرابات الهادئة" $\{\varepsilon_t\}$

<u>**الحل**</u>: حيث إن عملية $\{\varepsilon_t\}$ اضطرابات هادئة فإن:

$$E(\varepsilon_t) = 0\ ;\ \ t = 0, \pm 1, \pm 2, \dots..$$
$$V(\varepsilon_t) = \sigma^2\ ;\ \ t = 0, \pm 1, \pm 2, \dots..$$
$$\gamma(k) = Cov(\varepsilon_t, \varepsilon_{t-k})\ \ ;\ \ k \neq 0; t = 0, \pm 1, \pm 2$$

وبالتالي فإن: $\rho(k) = \frac{\gamma(k)}{\gamma(0)} = 0\ ,\ \ k \neq 0$ أي أن: $\rho(k) = \begin{cases} 1 & , k = 0 \\ 0 & , k \neq 0 \end{cases}$

<u>**مثال (7):**</u>

إذا كانت السلسلة y_t تتبع النموذج $y_t = \beta_0 + \beta_1 t + \varepsilon_t$ حيث ε_t عملية اضطرابات هادئة، أوجد دالة الارتباط الذاتي للسلسلةy_t

<u>**الحــل:**</u>

$$V(y_t) = V(\beta_0 + \beta_1 t + \varepsilon_t) = V(\varepsilon_t) = \sigma^2$$

وذلك لأن الدالة $f(t) = \beta_0 + \beta_1 t$ هي دالة محددة (غير عشوائية) Deterministi أي غير عشوائية.

$$y(s,t) = Cov(\beta_0 + \beta_1 s + \varepsilon_s , \beta_0 + \beta_1 t + \varepsilon_t) = 0 \ , \qquad s \neq t$$

ومن ثم فإن:

$$\rho(k) = \begin{cases} 1 & , k = 0 \\ 0 & , k \neq 0 \end{cases}$$

وتجدر الإشارة هنا إلى حقيقة في غاية الأهمية وهي أن دالة الارتباط الذاتي لهذا النموذج تساوي الصفر بدءًا من الفجوة الزمنية الأولى، أي أن هناك انقطاع تام لهذه الدالة على الرغم من عدم سكون هذه السلسلة، حيث إن لها اتجاه عام خطي بالزيادة أو النقصان على عكس ما قد نرى في الفصول القادمة عند التعامل مع نماذج ARIMA. وفي الواقع أنه لا يوجد تعارض بالمرة كما سنرى في الفصول القادمة، حيث إن النموذج y_t في هذا المثال الذي بين أيدينا ليس عشوائيًا nonstochastic بل هو نموذج محدد Deterministic كما أوضحنا سابقًا في الباب الأول، وبالتالي فعدم وجود ارتباط ذاتي بين مشاهدات السلسلة يبدو أمرًا منطقيًا. وقد ذكرنا هذا النموذج في هذا الباب لأنه عادة ما يحدث لبس للطالب أو الباحث الذي ليس لديه الخبرة والدراية الكافية بموضوعات النماذج المحددة والعشوائية وعلاقتها بالسكون وعلاقة هذا بالأخير بدالة الارتباط الذاتي. وبالطبع ما يقال عن النموذج y_t في هذا المثال يقال عن كل النماذج المحددة.

<u>مثال (8):</u> أوجد دالة الارتباط الذاتي للعملية $\{y_t\}$ في المثال (5)

<u>الحل:</u>

عند حل هذا المثال وجدنا أن:

$$\gamma(k) = \begin{cases} \sigma^2(1+\theta^2) & , k = 0 \\ -\theta\sigma^2 & , k = 1 \\ 0 & , k \geq 2 \end{cases}$$

وبالتالي فإن:

$$\rho(k) = \begin{cases} 1 & ; k = 0 \\ -\frac{\theta}{1+\theta^2} & ; k = 1 \\ 0 & ; k \geq 2 \end{cases}$$

3.1 تقدير دالة الارتباط الذاتي:

أوضحنا سابقاً أهمية وضع شروط السكون على العملية العشوائية التي ولدت السلسلة المرصودة (المتاحة) وأهمها تخفيض عدد المعالم الرئيسية (عزوم الدرجة الأولي والثانية وسهولة تفسيرها وإمكانية تقديرها باستخدام مشاهدات السلسلة المتاحة $y_1, y_2, \dots\dots, y_n$ وبناء على هذه التقديرات يمكن تقدير دالة الارتباط الذاتي للعملية العشوائية الساكنة بأحد التقديرين الآتيين:

$$r(k) = \hat{\rho}(k) = \frac{\sum_{t=1}^{n-k}(y_1 - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{n}(y_1 - \bar{y})^2}$$

$$r_0(k) = \tilde{p}(k) = \frac{\frac{1}{n-k}\sum_{t=1}^{n-k}(y_1-\bar{y})(y_{t+k}-\bar{y})}{\frac{1}{n}\sum_{t=1}^{n}(y_1-\bar{y})^2}$$

وفي الحقيقة أن هذين التقديرين متحيزان biased ، ولذلك فليس هناك أية أفضلية لإحداهما على الآخر، وعادة ما يستخدم التقدير الأول $r(k)$ لتقدير دالة الارتباط الذاتي، وهذا التقدير هو الذي سنستخدمه بالفعل في هذا الكتاب. ويمكن إثبات أنه إذا كانت العملية العشوائية $\{y_t\}$ ساكنة وخطية وأن العزم الرابع $E(Y_1^4)$ محدود فإن تقدير دالة الارتباط الذاتي $r(k)$ يتبع تقاربياً (إذا كانت n كبيرة) توزيع معتاد (معتدل) Normal متوسطه $\rho(k)$ وله تباين معين معروف يعتمد على $\rho(k)$ أيضاً. ومن ثم يمكن إجراء الاختبارات الإحصائية الخاصة بمعنوية significance الارتباطات الذاتية المختلفة.

والحالة الخاصة الهامة إذا كانت العملية العشوائية موضع الدراسة عملية "اضطرابات هادئة فإن تباين $r(k)$ يأخذ الصورة البسيطة الآتية: $V[r(k)] \approx \frac{1}{n}$

ومن ثم يمكن اختبار معنوية الارتباط الذاتي في هذه الحالة بشكل تقريبي كما سنرى عند تشخيص نماذج ARIMA في الباب الرابع. والمثال الآتي يوضح كيفية حساب معاملات الارتباط الذاتي:

<u>مثال (9):</u> تمثل البيانات الآتية عدد الوحدات المباعة بالمائة سنويًا من إحدى السلع في أحد المحلات الكبرى :

السنة	1992	1993	1994	1995	1996	1997	1998	1999
عدد الوحدات المباعة	1	3	2	4	3	2	3	2

احسب معاملات الارتباط الذاتي وارسم دالة الارتباط المقدرة.

<u>الحل:</u>

$$\bar{y} = \frac{20}{8} = 2.5 \ ; \ \sum_{t=1}^{8}(y_t-\bar{y})^2 = 6$$

$$r(1) = \frac{\sum_{t=1}^{7}(y_1-\bar{y})(y_{t+l}-\bar{y})}{6}$$

$$r(1) = \frac{1}{6}[(y_1-\bar{y})(y_2-\bar{y}) + (y_2-\bar{y})(y_3-\bar{y}) + (y_3-\bar{y})(y_4-\bar{y}) + (y_4-\bar{y})(y_5-\bar{y}) + (y_5-\bar{y})(y_6-\bar{y}) + (y_6-\bar{y})(y_7-\bar{y}) + (y_7-\bar{y})(y_8-\bar{y})] = -0.29$$

$$r(2) = \frac{1}{6}[(y_1-\bar{y})(y_3-\bar{y}) + (y_2-\bar{y})(y_4-\bar{y}) + (y_3-\bar{y})(y_5-\bar{y}) + (y_4-\bar{y})(y_6-\bar{y}) + (y_5-\bar{y})(y_7-\bar{y}) + (y_6-\bar{y})(y_8-\bar{y})] = 0.17$$

وبالمثل يمكن إثبات أنّ:

$$r(3) = -0.21; r'4) = -0.33\ ;\ \ r(5) = 0.21\ ; r(6) = -0.17;\ \ r(7) = 0.13$$

ومنه يمكن عرض دالة الارتباط الذاتي المقدرة في الشكل (3).

مثال (10): تمثل البيانات الآتية متوسط النسبة المئوية للسنوية للرطوبة في إحدى المدن:

السنة	1990	1991	1992	1993	1994
عدد الوحدات بالمائة	20	30	10	20	20

ارسم دالة الارتباط الذاتي لهذه البيانات.

الحل:

$$\bar{y} = \frac{100}{58} = 20\ \ ;\ \ \sum_{t=1}^{5}(y_t - \bar{y})^2 = 200$$

يمكن بسهولة إثبات أنّ: $r(1) = -\frac{1}{2}\ \ ;\ \ r(2) = r(3) = r(4) = 0$

ويمكن رسم هذه الدالة في الشكل (4).

ويلاحظ أن $r(\mathrm{K})$ تنقطع فجأة بعد الفجوة الزمنية الأولى.

2. دالة الارتباط الذاتي الجزئي Partial Autocorrelation Function

تلعب دالة الارتباط الذاتي الجزئي دوراً لا يقل أهمية عن دور دالة الارتباط الذاتي $\rho(k)-$ والتي سبق تقديمها في المبحث السابق في التعرف على النموذج الملائم للبيانات الزمنية المرصودة في منهجية بوكس وجينكنز. وقبل تعريف هذه الدالة ودراسة خصائصها في السلاسل الزمنية قد يكون من الأفضل أن نستهل هذا المبحث بمقدمة عن مفهوم الارتباط الجزئي بصفة عامة في مجالات الانحدار المألوفة لدى القارئ ثم ننتقل إلى تعميم هذا المفهوم في مجالات السلاسل الزمنية ودراسة خصائص دالة الارتباط الذاتي الجزئي وأهم طرق تقديرها باستخدام بيانات سلسلة زمنية متاحة.

1.2 مقدمة:

افترض أن $\mathrm{X, Z, W}$ ثلاثة متغيرات عشوائية لهم دالة كثافة احتمال مشترك $f(x, z, w)$. من المعروف في موضوعات الإحصاء بصفة عامة – وفي موضوعات الانحدار بصفة خاصة – أن معامل الارتباط الخطي بين المتغيرين $\mathrm{X, Z}$ يعرف في: $\rho(X, Z) = \rho_{X,Z} = \frac{Cov(X,Z)}{\sqrt{Var(X)Var(Z)}}$

ويقيس المعامل $\rho_{X,Z}$ درجة الاعتماد الخطي (الكلي) بين المتغيرين X, Z أي قوة الارتباط الخطي بينهما إذا كانت العلاقة بينهما على الشكل التالي (بافتراض أن X هو المتغير التابع):

$$E(X\ /\ Z) = \beta_0 + \beta_1 Z$$

ويمثل بسط معامل الارتباط الخطي التغاير بين المتغيرين والذي يمكن الحصول عليه من دالة كثافة الاحتمال المشتركة للمتغيرين X, Z, وذلك بإيجاد $E(X-\mu_X)(Z-\mu_Z)$ ، بينما يمثل المقام الجذر التربيعي لحاصل ضرب تباين المتغيرين. ويمكن الحصول على تباين المتغير X بإيجاد $E(X-\mu_X)^2$ باستخدام دالة كثافة الاحتمال الهامشي للمتغير X ، بينما يمكن الحصول على تباين المتغير Z بإيجاد $E(Z-\mu_Z)^2$ باستخدام دالة كثافة الاحتمال الهامشي للمتغير Z. إذا كان $\rho_{X,Z} > 0$ فإن هذا يعني أن القيم الكبرى للمتغيرين تميل أن تحدث معاً، كما أن القيم الصغرى لهما تميل أيضاً أن تحدث معاً. أما إذا كان $\rho_{X,Z} < 0$ فإن هذا يعني أن القيم الكبرى لأحد المتغيرين تميل أن تحدث مع القيم الصغرى للمتغير الآخر.

ولكن من جهة أخرى قد يكون هناك علاقة بين كل من المتغيرين X, Z بالمتغير الأخر W ، وفي هذه الحالة فإن $\rho_{X,Z}$ لا يعبر عن صافي العلاقة بين المتغيرين X, Z وإنما تعتمد قيمة هذا المعامل – بالإضافة إلى العلاقة بين X, Z – إلى مدى ارتباط كل من هذين المتغيرين بالمتغير الثالث W . فتغير المتغير الثالث W يساهم في تغير كل من المتغيرين X, Z ومن ثم يتأثر معامل الارتباط $\rho_{X,Z}$ بهذا التغير. وفي كثير من الأحيان قد يكون من المرغوب فيه البحث في صافي العلاقة المتغيرين X, Z بعد حذف تأثير المتغير W على هذين المتغيرين أي بافتراض ثبات المتغير W . ولإيجاد مثل هذا الارتباط – والذي يرمز له عادة بالرمز $\rho_{X,Z,W}$– يجب أولاً إيجاد التوزيع الشرطي $f(x,z|w)$ واستخدام هذا التوزيع لإيجاد معامل الارتباط الشرطي.

$$\rho_{X,Z,W} = \frac{Cov(X,Z|W)}{\sqrt{Var(X|W).Var(Z|W)}} = \frac{E[X-E(X|W)][Z-E(Z|W)]}{\sqrt{E[X-E(X|W)]^2.E[Z-E(Z|W)]^2}}$$

والسبب في اقتراح الصيغة السابقة لقياس الارتباط الجزئي بين المتغيرين X, Z يعود إلى أن المتغير العشوائي $[X-E(X|W)]$ يمثل المتغير العشوائي X بعد حذف تأثير المتغير W، كما أن المتغير العشوائي $[Z-E(Z|W)]$ يمثل المتغير العشوائيZ بعد حذف تأثير المتغير W. ومن ثم فإن معامل الارتباط بين المتغير $[X-E(X|W)]$ والمتغير $[Z-E(Z|W)]$ والمعرف بالصيغة السابقة – يكون اقتراح منطقي لقياس درجة الارتباط الجزئي بين المتغيرين X, Z بعد حذف تأثير المتغير W.

نظرية (1):

إذا كانت المتغيرات العشوائية X, Z, W تتبع توزيع معتاد ثلاث trivariate normal distribution فإن:

$$\rho_{X,Z,W} = \frac{\rho_{X,Z} - (\rho_{X,W})(\rho_{,Z,W})}{\sqrt{(1-\rho^2)(1-\rho^2)}}$$

ولن نتعرض لإثبات هذه النظرية هنا لأننا سنثبت الوجه الآخر لهذه النظرية في مجال السلاسل الزمنية في المبحث التالي.

2.2 ماهية الارتباط الذاتي الجزئي:

في موضوعات السلاسل الزمنية تحظى دراسة معامل الارتباط الذاتي الجزئي عند الفجوة الزمنية الثانية بأهمية خاصة، حيث يقيس هذا المعامل درجة الارتباط الخطي بين متغيرين يبعدان عن بعضهما البعض وحدتان زمنيتان بعد حذف تأثير المتغير الذي يقع بينهما، أي بافتراض ثبات هذا المتغير. فهذا المعامل يقيس قوة العلاقة الخطية بين المتغيرين Y Y بعد حذف تأثير المتغير Y2 ، ويقيس قوة العلاقة الخطية بين المتغيرين Y_4, Y_2 بعد حذف تأثير المتغير Y_3، ... وهكذا. وبصفة عامة يقيس هذا المعامل قوة الارتباط الخطي بين المتغيرين Y_t, Y_{t-2} بعد حذف تأثير المتغير الذي يقع بينهما وهو Y_{t-1}، أي بافتراض ثبات هذا المتغير ويرمز لهذا المعامل عادة بالرمز ϕ_{22} ويمكن تفسيره على أنه معامل الارتباط الخطي بين المتغير $[Y_t - E(Y_t|Y_{t-1})]$ والمتغير $[Y_{t-2} - E(Y_{t-2}|Y_{t-1})]$، ويمكن حسابه باستخدام دالتي الاحتمال الشرطي $f(Y_t, Y_{t-2}|Y_{t-1})$ و $f(Y_1|Y_{t-1})$

<u>نظرية (2):</u>

إذا كانت العملية $\{Y_t\}$ ساكنة وكانت المتغيرات Y_t, Y_{t-1}, Y_{t-2} تتبع توزيعاً معتاداً (معتدلاً) ثلاثياً فإن:

حيث إن المتغيرات الثلاثة تتبع توزيع معتاد متعدِّد فإن: $\phi_{22} = \frac{\rho(2)-\rho^2(1)}{1-\rho^2(1)}$

البرهان:

حيث أن المتغيرات الثلاثة تتبع توزيع معتاد متعدِّد فإن:

$$E(Y_{t-2}|Y_{t-1}) = \mu + \rho(1)[Y_{t-1} - \mu] \qquad (1)$$
$$E(Y_t|Y_{t-1}) = \mu + \rho(1)[Y_{t-1} - \mu] \qquad (2)$$
$$Var(Y_{t-2}|Y_{t-1}) = \sigma^2[1 - \rho^2(1)] \qquad (3)$$
$$Var(Y_t|Y_{t-1}) = \sigma^2[1 - \rho^2(1)] \qquad (4)$$
$$\phi_{22} = Corr\{[Y_t - E(Y_t|Y_{t-1})] \ ,[Y_{t-2} - E(Y_{t-2}|Y_{t-1})]\}$$
$$= \frac{E[Y_t - E(Y_t|Y_{t-1})]\ [Y_{t-2} - E(Y_{t-2}|Y_{t-1})]}{\sqrt{Var(Y_t|Y_{t-1}).E(Y_{t-2}|Y_{t-1})}} \qquad (5)$$

بالتعويض من (1) و (2) في بسط المعامل ϕ_{22} نصل إلى:

$$\text{البسط} = E[Y_1 - \mu - \rho(1)(Y_{t-2} - \mu)][Y_{t-2} - \mu - \rho(1)(Y_{t-1} - \mu)]$$
$$= E[Y_1 - \mu][Y_{t-2} - \mu] - \rho(1)E(Y_1 - \mu)(Y_{t-1} - \mu)$$
$$-\rho(1)E(Y_{t-1} - \mu)(Y_{t-2} - \mu) + \rho^2 E(Y_1 - \mu)^2$$

$$= \gamma(2) - \rho(1)\gamma(1) - \rho(1)\gamma(1) + \frac{\rho(1)\gamma(0)\gamma(1)}{\gamma(0)} = \gamma(2) - \rho(1)\gamma(1) \qquad (6)$$

بالتعويض من (1) و (2) في مقام المعامل ϕ_{22} في المعادلة (5) نصل إلى:

$$\text{المقام} = \gamma(0)[1 - \rho^2(1)] \qquad (7)$$

بقسمة (6) على (7) نصل إلى: $\phi_{22} = \frac{[\gamma(2)-\rho(1)\gamma(1)]}{\gamma(0)[1-\rho^2(1)]} = \frac{\rho(2)-\rho^2(1)}{1-\rho^2(1)}$

وهو المطلوب إثباته.

وتجدر الإشارة إلى أنه يمكن بسهولة إثبات النظرية (2) بالتعويض عن النظرية (1) مباشرة وذلك بوضع:

$$Y_{t-2} = X\ ;\ Y_t = Z\ ;\ \ Y_{t-1} = W$$

تعريف:

يعرف معامل الارتباط الجزئي للعملية الساكنة $\{Y_t\}$ عند الفجوة الزمنية k بأنه معامل الارتباط الخطي بين المتغيرين Y_t, Y_{t-k} بعد حذف تأثير المتغيرات الت تقع بينهما وهي $Y_{t-1}, Y_{t-2}, \dots Y_{t-k+1}$

ويرمز عادة لمعامل الارتباط الجزئي عند الفجوة k بالرمز ϕ_{22} ويمكن تفسيره على أنه معامل الارتباط الخطي بين المتغير $[Y_t - E(Y_t|Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1})]$ والمتغير $[Y_{t-k} - E(Y_t|Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1})]$. ويمكن حساب معامل الارتباط الذاتي باستخدام دوال الاحتمال الشرطي المناسبة لجميع قيم $k = 1, 2, \dots$

تعريف:

تعرف دالة الارتباط الذاتي الجزئي بأنها علاقة دالية بين معاملات الارتباط الجزئي ϕ_{kk} والفجوة الزمنيةk.

وتعرض دالة الارتباط الجزئي – شأنها في ذلك شأن دالة الارتباط الذاتي – عادة في شكل رياضي وأحيانًا في شكل جدولي أو بياني (شعراوي، 2005 صفحة 125).

3.2 خصائص دالة الارتباط الذاتي الجزئي:

تتصف دالة الارتباط بعدة خصائص هامة نذكر منها ما يلي:

1- معامل الارتباط الذاتي الجزئي عند الفجوة الزمنية صفر يساوي واحد، أي أن: $\phi_{00} = 4$لأي عملية ساكنة.

2- قيمة ϕ_{00} تقع دائمًا على الفترة المغلقة $[-1,1]$

3- معامل الارتباط الذاتي الجزئي عند الفجوة الزمنية الأولي دائمًا يساوي معامل الارتباط الذاتي عند الفجوة الزمنية الأولى، أي أن$\phi_{11} = \rho(1)$ وذلك لعدم وجود متغيرات بين المتغيرين Y_t, Y_{t-1}

4- إذا كان $\phi_{kk} = 0$ فهذا يعني أنه لا توجد علاقة خطية جزئية بين أي متغيرين الفاصل الزمني بينهما k وحدة، ولكن بالطبع قد توجد علاقة جزئية غير خطية بينهما.

وتأخذ الدالة ϕ_{kk} في التحليل الحديث أشكالاً قريبة الشبه من أشكال دالة الارتباط الذاتي $\rho(k)$ ، فتارة تتلاشى ببطء، وتارة تتلاشي بسرعة في صورة أسية وتارة تقترب تدريجيًا من الصفر في شكل موجات تحاكي دالة الجيب أو في شكل توليفة من الدوال الأسية، وتارة تنقطع كلية بعد عدد معين من الفجوات الزمنية. وتلعب دالة الارتباط الذاتي الجزئي دوراً لا يقل أهمية عن دور دالة الارتباط الذاتي، فتستخدم لاختبار سكون السلسلة بجانب الطرق الأخرى، وهى أحد الأدوات الرئيسية الت وظفت بواسطة بوكس وجينكنز للتعرف على النموذج المبدئي وتشخيص هذا النموذج من أجل تحسينه أو تطويره وسنتعرض للدور الذي تلعبه هذه الدالة بالتفصيل عند تقديم منهجية بوكس وجينكنز في الباب الرابع.

4.2 تقدير دالة الارتباط الذاتي الجزئي:

قدم الفكر الخاص بالسلاسل الزمنية أساليب عديدة لتقدير دالة الارتباط الذاتي الجزئي ويعتبر أسلوب أو نظام يوول – والكر من أهم هذه الأساليب على الإطلاق. ونظرا لأهمية هذا النظام والدور الهام الذي يلعبه في منهجية بوكس وجينكنز فقد خصَّصنا المبحث القادم بالكامل لعرض هذا النظام بالتفصيل، بينما نقدم في هذا المبحث أسلوبين آخرين لتقدير الارتباط الذاتي الجزئي.

الأسلوب الأول:

لدراسة الارتباط الجزئي بين المتغيرين Y_t, Y_{t-k} يفترض هذا الأسلوب أن علاقة الانحدار بين المتغير Y_{t-k} والمتغيرات التي تقع بينه وبين المتغير Y_t على الصورة الخطية الآتية:

$$Y_{t-k} = \beta_0 + \beta_1 Y_{t-k+1} + \beta_2 Y_{t-k+2} + \cdots + \beta_{k-1} Y_{t-1} + \varepsilon_{t-k}\ ; k = 2,3, \dots n-1 \qquad (2.3.1)$$

وبالتالي فإن المتغير العشوائي:

$$\varepsilon_{t-k} = Y_{t-k} - (\beta_0 + \beta_1 Y_{t-k+1} + \beta_2 Y_{t-k+2} + \cdots + \beta_{k-1} Y_{t-1})$$

يمثل المتغير Y_{t-k} بعد حذف تأثير المتغيرات التي تقع بينه وبين المتغير Y_t . وبالمثل يفترض هذا الأسلوب أن علاقة الانحدار بين المتغير Y_t والمتغيرات التي تقع بينه وبين المتغير Y_{t-k} على الصورة الخطية:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-k+1} + \alpha_2 Y_{t-k+2} + \cdots + \alpha_{k-1} Y_{t-1} + \varepsilon_{t-k}\ ; k = 2,3, \dots n-1 \qquad (2.3.2)$$

ومن ثم فإن المتغير العشوائي: $e_t = Y_t - (\alpha_0 + \alpha_1 Y_{t-k+1} + \alpha_2 Y_{t-k+2} + \cdots + \alpha_{k-1} Y_{t-1})$

يمثل المتغير Y_t بعد حذف المتغيرات التي تقع بينه وبين المتغير Y_{t-k} وبناءً على ذلك يمكن تقدير ϕ_{kk} عن طريق إجراء الخطوات الآتية:

1- نضع $k = 2$ ونجري الانحدار (2.3.1) ونحصل منه على البواقي $\hat{\varepsilon}_t$، ثم نجري الانحدار (2.3.2) ونحصل منه على البواقي $\hat{e}_t$.

2- نحسب معامل بيرسون للارتباط الخطي بين قيم القيم $\hat{\varepsilon}_t$ والقيم $\hat{e}_t$. هذا المعامل يعطي تقدير مناسب لمعامل الارتباط الجزئي ϕ_{22} ، ويرمز له عادة بالرمز $\hat{\phi}_{22}$.

3- نضع $k = 3$ ونكرر الخطوتين السابقتين ونحصل على تقدير لمعامل الارتباط الذاتي الجزئي ϕ_{33} وليكن $\hat{\phi}_{33}$.

4- نكرر الخطوتين 2,1 لجميع القيم الأخرى للفجوة الزمنية k ، وبالتالي يمكن الحصول على التقديرات . $\hat{\phi}_{22}, \hat{\phi}_{33}$.

الأسلوب الثاني:

بالرغم من سهولة الأسلوب الأول، إلا إنه يحتاج إلى توفيق معادلتي انحدار مختلفتين لتقدير كل معامل. ومن ثم لا بد من توفيق عدد من معادلات الانحدار يساوي ضعف عدد معاملات الارتباط الذاتي الجزئي المطلوب تقديرها. الأسلوب الثاني يوفر نصف هذا العدد وذلك بتوفيق مجموعة معادلات الانحدارات الآتية والحصول على آخر تقدير في كل معادلة ليمثل التقدير المطلوب. (شعراوي، 2005 صفحة 127)

$$
\begin{array}{ll}
1. & Y_t = \phi_{21}Y_{t-1} + \phi_{22}Y_{t-2} + \varepsilon_t \\
2. & Y_t = \phi_{31}Y_{t-1} + \phi_{32}Y_{t-2} + \phi_{33}Y_{t-3} + \varepsilon_t \\
\vdots & \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\
k. & Y_t = \phi_{k1}Y_{t-1} + \phi_{k2}Y_{t-2} + \cdots . + \phi_{kk}Y_{t-k} + \varepsilon_t
\end{array}
$$

وقبل أن نختتم هذا المبحث تجدر الإشارة بالقول بأنه قد لا يكون هناك داع لتقدير المعامل ϕ_{11} بشكل مستقل حيث إن هذا المعامل يساوي بالتعريف معامل الارتباط الذاتي $\rho(1)$ ، لأنه ليس هناك أي متغيرات تقع بين Y_t, Y_{t-1} ، ومن ثم يمكن تقدير ϕ_{11} كما يلي: $\hat{\phi}_{11} = \hat{\rho}(1) = r(1)$.

المحور الثامن:

نموذج تمهيد الأسي للتنبؤ بالسلاسل الزمنية.

1.التمهيد الأسى Exponential Smoothing

طريقة المتوسطات المتحركة البسيطة السابق ذكرها من الطرق الخاصة جدًا والتي تفيد فقط إذا كانت بيانات السلسلة يغلب عليها الطابع العشوائي، أي إذا كانت البيانات تتأرجح بشكل غير نمطي حول متوسط ثابت يمثل مستوى السلسلة على الفترة الزمنية موضع الدراسة (والتر، 1992 صفحة 235). ولكن في الكثير من التطبيقات قد يتغير متوسط الظاهرة ببطء على الفترة الزمنية موضع الدراسة، وفي مثل هذه الحالات قد يكون من المنطقي إعطاء وزن أكبر لأحدث مشاهدة عند التنبؤ وأوزانًا تتناقص بشكلٍ أو بآخر بزيادة عمر المشاهدة أي بزيادة الفاصل الزمني بين زمن المشاهدة والزمن الذي يراد التنبؤ عنده. وبالطبع يوجد العديد من الدوال الرياضية التي تعكس مفهوم تناقص الأوزان أو الأهميات بزيادة عمر المشاهدة، إلاّ أن أهم هذه الدوال ما يعرف بالدوال الأسية والتي وجدت أرضية خصبة وممهدة ليس في أدبيات السلاسل الزمنية التقليدية فحسب بل في الأدبيات الحديثة أيضًا. وتعتمد فكرة هذه الدوال على إعطاء وزن ترجيح كبير لأحدث مشاهدة عند الزمن الذي يراد التنبؤ عنده ثم إعطاء أوزان ترجيحية تتناقص بشكل أسي مع زيادة الفاصل الزمني بين زمن التنبؤ وزمن المشاهدة.

إذا افترضنا أننا نقف عند نقطة أصل معينة t ونريد التنبؤ بقيمة الظاهرة عند الفترة الزمنية $t+1$ وأن هذا التنبؤ يرمز له بالرمز $\hat{y}_1(1)$، فإن نموذج التمهيد الأسي يعرف على الصورة:

$$\hat{y}_1(1) = \frac{1}{c}\left[y_t + wy_{t-1} + w^2 y_{t-2} + \cdots w^{t-1} y_t\right] \quad \ldots\ldots\ldots\ldots\ldots\ldots (1.6.8)$$

حيث:

$$0 < w < 1 \ ; \ \ c = \sum_{i=0}^{t-1} w^t = \frac{(1-w^t)}{(1-w)}$$

ويسمي العدد w بمعامل التناقص discount coefficient، ويشير الثابت c ، إلى مجموع أوزان الترجيح التي تجعل من النموذج (1.6.8) متوسط حقيقي. ويعرف النموذج (1.6.8) عادة في أدبيات السلاسل الزمنية بنموذج المتوسطات المتحركة المرجح آسيًا exponentially weighted moving average model ويشار إليه عادة بالحروف EWMA. وتعتمد قيمة معامل التناقص w على سرعة التغير في مستوي السلسلة وعادة ما يكون $0.7 < w < 0.95$

إذا كانت t كبيرة فإن $w^t \to 0$ وفي هذه الحالة يمكن إعادة كتابة النموذج (1.6.8) على الصورة:

$$\hat{y}_t(1) = (1-w)y_t + (1-w)wy_{t-1} + (1-w)w^2 y_{t-2} + \cdots \ldots. \qquad (1.6.9)$$

وتجدر الإشارة إلى أن $\hat{y}_t(1)$ لا يمثل بالفعل متوسطًا حقيقيًا لأن مجموع أوزان الترجيح $(1-w),(1-w)w,(1-w)w^2\ldots$.. يساوي الواحد الصحيح. كما تجدر الإشارة إلى أن نظام الترجيح (1.6.9) يمكن تعديله باختيار قيم مختلفة للمعامل w ، فإذا كانت w كبيرة فإن وزن الترجيح الذي يعطى المشاهدة الحالية لا يكون صغيرًا وأوزان الترجيح المتتابعة تتناقص ببطء، أما إذا كانت w صغيرة فإن وزن الترجيح الذي يعطى للمشاهدة الحالية يكون كبيرًا وأوزان الترجيح المتتابعة تتناقص بسرعة. فعلى سبيل المثال إذا كانت $w=0.9$ فإن معاملات المشاهدات $y_t,y_{t-1},y_{t-2},\ldots$ تكون على الترتيب $0.1,0.09,0.081,\ldots$ أي أن تأثير المشاهدات يتضاءل ببطء، أما إذا كانت $w=0.1$ فإن معاملات نفس المشاهدات $y_t,y_{t-1},y_{t-2},\ldots$ لاستكون على الترتيب $0.9,0.09,0.009$ أي أن تأثير المشاهدات يتضاءل بسرعة بدءًا من المشاهدة الحالية.

ويمكن كتابة التنبؤ عند الفترة الزمنية السابقة $(t-1)$ من الصورة (1.6.9) كما يلي:

$$\hat{y}_{t-1}(1)=(1-w)y_{t-1}+(1-w)wy_{t-2}+(1-w)w^2y_{t-3}+\cdots \quad (1.6.10)$$

وبالتعويض من (1.6.10) في (1.6.9) نصل إلى الصيغة التكرارية (التتابعية) Recurrence Formula الآتية:

$$\hat{y}_t(l)=(1-w)y_t+w\hat{y}_{t-I}(I) \qquad (1.6.11)$$

وتوضح الصيغة (1.6.11) أن التنبؤ الحديث عند الزمن t يساوي الوسط الحسابي المرجح للتنبؤ السابق له مباشرة $\hat{y}_{t-I}(I)$ والمشاهدة الحديثة y_t ووزني الترجيح هما w و $(1-w)$ على الترتيب وتستأثر المشاهدة الأحدث بالنصيب الأكبر في هذه العلاقة إذا كانت قيمة w صغيرة، بينما تكون مساهمة هذه المشاهدة صغيرة إذا كانت قيمة w كبيرة. وتمكن الصيغة (1.6.11) من حساب التنبؤات للسلسلة بشكل تتابعي كالآتي:

$$\hat{y}_1(I)=(1-w)y_1+w\hat{y}_0(1) \qquad (1)$$
$$\hat{y}_2(I)=(1-w)y_2+w\hat{y}_1(1) \qquad (2)$$
$$\hat{y}_3(1)=(1-w)y_3+w\hat{y}_2(1) \qquad (3)$$
$$\vdots$$

بالتعويض من المعادلة (1) في المعادلة (2):

$$\hat{y}_2(1)=(1-w)y_2+w(1-w)y_1+w^2\hat{y}_0(1) \qquad (4)$$

بالتعويض من المعادلة (4) في المعادلة (3):

$$\hat{y}_3(1)=(1-w)y_3+w(1-w)y_2+w^2(1-w)y_1+w^3\hat{y}_0(1)$$

بالاستمرار في هذه العملية نصل إلى:

$$\hat{y}_t(1)=[(1-w)y_t+w(1-w)y_{t-1}+w^2(1-w)y_{t-2}+\cdots+w^{t-1}y_1]+w^1\hat{y}_0(1) \qquad (1.6.12)$$

وتوضح الصورة (1.6.12) أن التنبؤ عند الفترة الزمنية t، يمكن التعبير عنه بدلالة مشاهدات السلسلة المتاحة $y_t,y_{t-1},\ldots.y_1$ والتنبؤ الابتدائي $\hat{y}_0(1)$. وتجدر الإشارة هنا إلى ملاحظتين أساسيتين حول الصيغة (1.6.12). الملاحظة الأولى أن معاملات شاهدات $y_t,y_{t-1},\ldots.y_1$ هي على الترتيب $(1-w),(1-w)$

$w)w, (1-w)w^2, \dots (1-w)w^{t-1}$ وأن هذه المعاملات تقيس مساهمة هذه المشاهدات في التنبؤ $\hat{y}_t(1)$. وكما هو واضح أن هذه المعاملات تعطي أهميات أكبر للمشاهدات الحديثة وتتناقص قيم هذه المعاملات بشكل أسي. والملاحظة الثانية أن المعامل w^t يقيس مساهمة التنبؤ المبدئي $\hat{y}_0(1)$ على التنبؤ $\hat{y}_1(1)$ ، ومن ثم يتضاءل تأثير التنبؤ المبدئي $\hat{y}_0(1)$ على التنبؤ $\hat{y}_1(1)$ إذا كان طول السلسلة n معقولاً. ولحساب التنبؤات عادة ما تستخدم الصيغة التكرارية (1.6.11) لسهولة تحديث التنبؤات حيث يكفي معرفة المشاهدة الحديثة والتنبؤ السابق مباشرة. ولحساب التنبؤات يجب معرفة القيمة الابتدائية $\hat{y}_0(1)$ ومعامل التناقص w . بالنسبة لاختيار القيمة الابتدائية يوجد العديد من الطرق لتقديرها أهمها:

1- استخدام الوسط الحسابي لقيم السلسلة $y_1, y_2, \dots, y_n$، وهذه الطريقة عادة ما تكون ملائمة إذا كان متوسط السلسلة يتغير ببطء على الفترة الزمنية موضع الدراسة.

2-ويفضل بعض الباحثين استخدام المشاهدة الأولي y_1 كتقدير للقيمة الابتدائية.

3- استخدام الوسط الحسابي لبعض المشاهدات الأولى لتقدير هذه القيمة.

للمزيد من التفاصيل حول اختيار القيمة الابتدائية يمكن للقارئ الرجوع إلى (1962) Brown أو إلى (1976) Montgomery and Johnson أو إلى Abraham and (1983) Ledolter أو إلى (1987) Bowerman and O'Connell

ويستأثر معامل التناقص w بأهمية خاصة في التأثير على التنبؤ $\hat{y}_t(1)$, ومن ثم يحظى هذا المعامل بأهمية خاصة عند اختياره. وهناك بعض الخطوط العامة التي يمكن الاسترشاد بها عند هذا الاختيار والخاصة بسرعة التقلبات التي تحدث في السلسلة. فإذا كانت السلسلة تتعرض للكثير من التقلبات غير المنتظمة فقد يكون من الأفضل استخدام قيمة كبيرة للمعامل w وذلك من أجل إعطاء وزن وأهمية للتنبؤ السابق $\hat{y}_{t-1}(1)$, أكبر من وزن المشاهدة الحديثة $\hat{y}_t(1)$، ويؤدي هذا إلى التخلص من الأخطاء العشوائية والحصول على تنبؤات مستقرة. أما إذا كانت السلسلة أكثر هدوءًا واستقرارًا أو يوجد تغير منتظم في نمط السلسلة فقد يكون من الأفضل اختيار قيمة صغيرة للمعامل w وذلك من أجل إعطاء وزن أكبر للمشاهدة الحديثة (شعراوي، 2005 صفحة 31). ولكن هذه الخطوط العامة لا تمكن بالطبع من اختيار قيمة دقيقة لهذا المعامل في التطبيقات العملية ولذلك عادة ما يتم اختيار هذا المعامل عن طريق المحاكاة simulation في مثل هذه التطبيقات حيث يتم توليد تنبؤات مختلفة للمعامل w ثم تقارن هذه التنبؤات بالقيم العملية للسلسلة $y_1, y_2, \dots, y_n$, لحساب الأخطاء المناظرة لكل قيمة من قيم المعامل w. بعد ذلك يتم حساب أحد المقاييس التي سبق دراستها لقياس حجم الأخطاء وليكن مجموع المربعات المناظر لكل قيمة من قيم w ، وتكون قيمة المعامل w المناسبة هي القيمة التي تجعل هذا المجموع أقل ما يمكن. وطريقة المحاكاة ليست الطريقة الوحيدة المعروفة لاختيار قيمة w ولكن هناك طرق أخرى أهمها ما

يعرف بطريقة التجربة والخطأ trial and error ، ولن نتعرض لدراسة هذه الطريقة هنا ولكن يمكن للقارئ الرجوع إلى Gaynor and Kirkpatrick (1994) للتعرف عليها والمثال الآتي يوضح كيفية استخدام طريقة التمهيد الأسي في التنبؤ .

مثال (4): البيانات الآتية تمثل عدد الأجهزة (بالمائة) المباعة التي سجلت شهريًا في دفاتر إحدى الشركات:

14 16 17 13 15 14 15 13 14 12 12 11 16

قدر القيمة الابتدائية $\hat{y}_0(1)$ باستخدام الوسط الحسابي لقيم السلسلة ثم استخدم هذه القيمة لإيجاد التنبؤات المناظرة مرة باستخدام $w = 0.5$ ومرة أخرى باستخدام $w = 0.9$، أي التنبؤات أفضل؟ اشرح سبب إجابتك.

الحـل:

$$\hat{y}_0(1) = \bar{y} = \frac{1}{13}[11 + 12 + 12 + \cdots + 16] = 14$$
$$\hat{y}_t(1) = (1 - w)y_t + w\hat{y}_{t-1}\ (1)$$

إذا كانت $w = 0.7$:

$$\hat{y}_1(1) = 0.3(y_1) + 0.7\hat{y}_0(1) = 0.3(11) + 0.7(14) = 13.1$$
$$\hat{y}_2(1) = 0.3(y_2) + 0.7\hat{y}_1(1) = 0.3(12) + 0.7(13.1) = 12.77$$
$$\hat{y}_3(1) = 0.3(y_3) + 0.7\hat{y}_2(1) = 0.3(12) + 0.7(12.77) = 12.539$$

وبالاستمرار في هذه العملية يمكن توليد التنبؤات الموضحة في جدول (1)

إذا كانت $w = 0.9$:

$$\hat{y}_1(1) = 0.1(y_1) + 0.9\hat{y}_0(1) = 0.1(11) + 0.9(14) = 13.7$$
$$\hat{y}_2(1) = 0.1(y_2) + 0.9\hat{y}_1(1) = 0.1(12) + 0.9(13.7) = 12.53$$
$$\hat{y}_3(1) = 0.1(y_3) + 0.9\hat{y}_2(1) = 0.1(12) + 0.9(13.53) = 13.377$$

وهكذا يمكن توليد باقي التنبؤات والموضحة في جدول (1)

جدول (1): القيم الفعلية والتنبؤات لبيانات المثال (4)

t	y_t	التنبـــــؤات		${e^2}_t = [y_t - \hat{y}_{t-1} - 1)]^2$	
		$w = 0.7$	$w = 0.9$	$w = 0.7$	$w = 0.9$
1	11	14	14	9.00	9.00
2	12	13.1	13.7	1.21	2.89
3	12	12.77	13.53	0.59	2.34
4	14	12.539	13.38	2.13	0.38
5	13	12.98	13.34	0.00	1.12

6	15	12.98	13.51	4.08	2.22
7	14	13.59	13.66	0.17	0.12
8	15	13.17	13.69	3.35	1.72
9	13	14.10	13.82	1.21	0.67
10	17	13.77	13.74	10.43	10.63
11	16	17.74	14.07	1.59	3.72
12	14	15.12	14.26	1.25	0.07
13	16	14.78	14.23	1.49	3.13
		15.15	14.41	–	–

بفحص نتائج الجدول (1) نجد أن مجموع مربعات الأخطاء المناظرة للقيمة $w = 0.7$ هو:

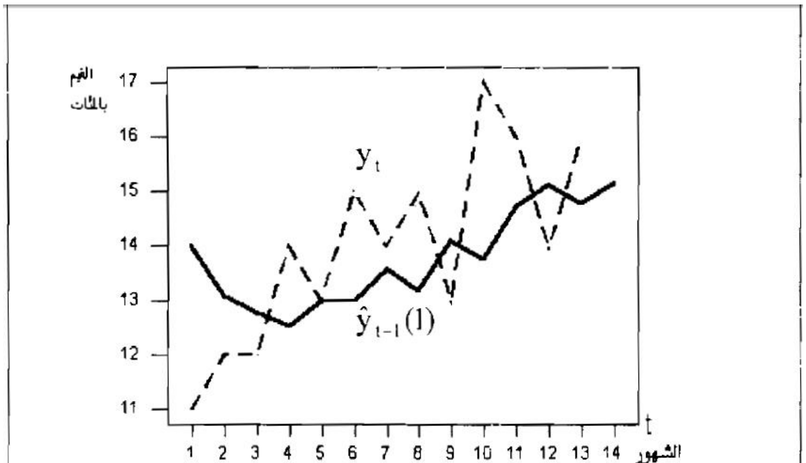
$$S\,(0.7) = 9 + 1.121 + 0.59 + \ldots + 1.4\,9 = 35.5$$

وبالمثل فإن مجموع مربعات الأخطاء المناظرة للقيمة $w = 0.9$ هو:

$$S(0.9) = 9 + 2.89 + 2.34 + \ldots + 3.13 = 37.01$$

حيث إن $S\,(0.7) < S(0.9)$ يمكن القول بأن التنبؤات المناظرة لقيمة $w = 0.7$ أفضل من التنبؤات المناظرة لقيمة $w = 0.9$، ويكون التنبؤ بعدد الأجهزة المباعة في الشهر التالي هو 1515 جهازا ويعرض شكل (7) القيم الفعلية والتنبؤات التي حصلنا عليها إذا كانت $w = 0.7$

شكل (7): القيم الفعلية والتنبؤات للمثال (4)



ويلاحظ من شكل (7) أن سلسلة التنبؤات أكثر تمهيدًا ونعومة من السلسلة الأصلية.

وقد أوجد نموذج المتوسطات المتحركة المرجح أسيًا لنفسه طريقا ممهدًا في الكثير من تطبيقات وأدبيات التنبؤ في مجالات المعرفة المختلفة خاصة الاقتصادية والإدارية منها وذلك لعدة أسباب، السبب الأول هو سهولة تطبيقه وتحديث تنبؤاته باستخدام التنبؤ السابق والمشاهدة الأحدث فقط، وهذا السبب هو أحد الأسباب الهامة لانتشار طريقة التمهيد الأسي في هذه المجالات خاصة عند التنبؤ بالعديد من السلاسل الزمنية كما هو الحال عند التنبؤ بمبيعات الآلاف من السلع المختلفة الموجودة في أحد المحلات العملاقة. السبب الثاني في هذا الانتشار يعود إلى آلية الطريقة الكاملة، فبمجرد برمجة الخطوات الرئيسية واختيار قيمة المعامل w يمكن الحصول على التنبؤات دون تدخل الإرادة الإنسانية، هذه الميزة بالذات هامة جدا لمستخدمي الإحصاء في مجالات المعرفة المختلفة والذين يبحثون دائما عن أساليب لا تتطلب مهارات وخبرات خاصة. السبب الثالث وراء انتشار هذه الطريقة هو سهولة برمجة خطواتها الرئيسية وانخفاض تكاليف استخدامها مقارنة باستخدام النماذج العشوائية الحديثة. السبب الرابع أن استخدام هذه الطريقة لا يتطلب توافر عدد كبير من المشاهدات ولذلك وجدت هذه الطريقة أرضًا خصبة في البلاد النامية حيث تعاني معظم هذه البلاد من عدم توافر الحد الأدنى من المشاهدات الضروري لاستخدام النماذج العشوائية الحديثة.

وتتعرض طريقة التمهيد الأسي للعديد من الانتقادات، أولها عدم وجود منهجية عامة للاختيار بين أنظمة الترجيح البديلة وعدم وجود طريقة عامة لتقويم نظام الترجيح المختار. والانتقاد الثاني الذي يوجه إلى هذه الطريقة أنها تعالج كل السلاسل الزمنية التي تنشأ في الواقع بطريقة واحدة ولذلك فهي قد تؤدي إلى تنبؤات غير صالحة، والانتقاد الثالث الذي يوجه إلى هذه الطريقة أنه ليس من السهل دائما اختيار المعامل w بدقة وعدم وجود طريقة وحيدة لاختيار هذا المعامل والتقدير المبدئي $\hat{y}_0(1)$ ، ولذلك فإن التنبؤات التي نحصل عليها من هذه الطريقة قد تختلف من باحث إلى آخر. الانتقاد الأخير والهام أن هذه الطريقة تعتبر من الطرق الخاصة وذلك لأنها تكون ملائمة فقط إذا كانت العملية العشوائية الكامنة في البيانات لها مواصفات خاصة. بصورة أكثر تحديدًا يمكن القول بأن هذه الطريقة تعمل بصورة جيدة إذا كانت العملية العشوائية تنتمي إلى فئة جزئية فقط من النماذج العشوائية الحديثة والتي لها مواصفات خاصة والتي تكون مفيدة في بعض المواقف وغير المفيدة في مواقف أكثر (شعراوي، 2005 صفحة 39).

قائمة المراجع

بخيت علي ,حسين .2009 *.القياس الاقتصادي* : s.l. .دار اليازوري.2009 ,

بخيت ,حسين علي etفتح الله ,سحر .2010 *.الاقتصاد القياسي* .عمان الأردن :دار اليازوري.2010 ,

بخيت ,حسين علي etفتح الله ,سحر .2006 *.القياس الاقتصادي* .عمان :دار اليازوري .pp .2006 , 145-144.

خالد زهدي ,خواجة *.السلاسل الزمنية* .بغداد العراق :المعهد العربي للتدريب والبحوث الاحصائية.

دومينيك ,سالفاتور .1983 *.الاحصاء والاقتصاد القياسي* .مصر :الدار الدولية للنشر والتوزيع.1983 ,

شاكر ,فارس عياد etفتاوي ,عزت .2006 *.مبادئ الإقتصاد القياسي والرياضي* .الفيوم :دار العلم للنشر والتوزيع.2006 ,

شعراوي ,سمير مصطفى .2005 *.مقدمة في التحليل الحديث للسلاسل الزمنية* : s.l. .مركز النشر العلمي , 2005.

شعراوي ,سمير مصطفى .2005 *.مقدمة في التحليل الحديث للسلاسل الزمنية* .المملكة العربية السعودية : مركز النشر العلمي.2005 ,

عبد القادر عطية ,عبد القادر محمد .2009 *.الحديث في الاقتصاد القياسي بين النظرية والتطبيق* . الاسكندرية :الدار الجامعية.2009 ,

عبد القادر عطية ,عبد القادر .2004 *.الحديث في القياس الاقتصادي بين النظرية والتطبيق* : s.l. .الدار الجامعية للطباعة والنشر.2004 ,

عدنان داوي ,محمد العذاري .2010 *.الاقتصاد القياسي نظرية وحلول* .عمان الأردن :دار جرير للنشر والتوزيع.2010 ,

كامل علاوي ,فاضل الفتلاوي .2014 *.القياس الاقتصادي :النظرية والتحليل* .الأردن :دار صفاء للطباعة والنشر والتوزيع.2014 ,

نظير مدكور ,أمال .2007 *.محاضرات في الاقتصاد القياسي* .فلسطين :كلية التجارة.2007 ,

.2007-2006 .—*محاضرات في الاقتصاد القياسي* : s.l. .جامعة الأزهر.2007-2006 ,

والتر ,فاندل .**1992**. *السلاسل الزمنية ونماذج بوكس جانكس* .المملكة العربية السعودية :دار المريخ للنشر , 1992.

1996. .*ـــــالسلاسل الزمنية ونماذج بوكس جينكنز* : .s.l. دار المريخ للنشر .1996 ,

قائمة المحتويات