المحاضرة الاولى :مقدمة في القياس الاقتصادي

اولا : مفاهيم عامة حول الاقتصاد القياسي :

يعتبر الاقتصاد القياسي احد فروع العلوم الاقتصادية الذي يهتم بالبحث التجريبي ،ويتميز باستخدامه للأساليب الكمية في تحليل الظواهر الاقتصادية من اجل اغراض متعددة مثل التنبؤ ،التحليل الاقتصادي ،وضع السياسات الاقتصادية و اتخاذ القرارات الاقتصادية المثلى و غيرها.

يعود استخدام مصطلح الاقتصاد القياسي لاول مرة سنة 1926 من قبل الاقتصادي RAGNER FRISCH ،غير ان البداية الحقيقية للاقتصاد القياسي كانت سنة 1930 اثناء تاسيس الجمعية الدولية للقياس الاقتصادي من طرف كل من H.MORE، H.SCHULTRZ ، R.STONE ، R.FRISCH. الامر الذي نتج عنه تاسيس اول مجلة اكاديمية خاصة بالقياس الاقتصادي تحت مسمى ECONOMITRICA سنة 1933 والتي لازالت قيد الدراسة و النشر الى غاية اليوم ،والاقتصاد القياسي هو مصطلح يوناني يتكون من ECPNOMIC (الاقتصاد) ،و METRIC وتعني القياس ،وتترجم كلمة ECONOMITRIC في الاوساط العلمية العربية **بالاقتصاد القياسي** ،غير ان التسمية الصحيحة لهذا الفرع من العلوم الاقتصادية هي **القياس الاقتصادي** وبشكل عام الاقتصاد القياسي هو علم يهتم بقياس العلاقات الاقتصادية و تكميمها (جعلها كمية) من خلال بيانات احصائية مستخدما الطرق الرياضية في ذلك.

ثانيا :تعريف الاقتصاد القياسي :

حسب **برونس هانسن** يعتبر افضل تعريف للقياس الاقتصادي هو ذلك التعريف الذي تم ارفاقه فريش في اول عدد من مجلة ECONOMITRICAسنة 1933 حين تم تعريف جمعية القياس الاقتصادي على انها مجتمع دولي اقيم للنهوض بالنظرية الاقتصادية فيما يتعلق بالرياضيات و الاحصاء هدفها الرئيسي هو توحيد النظرية الكمية و المنهج التجريبي مع المشكلات الاقتصادية.

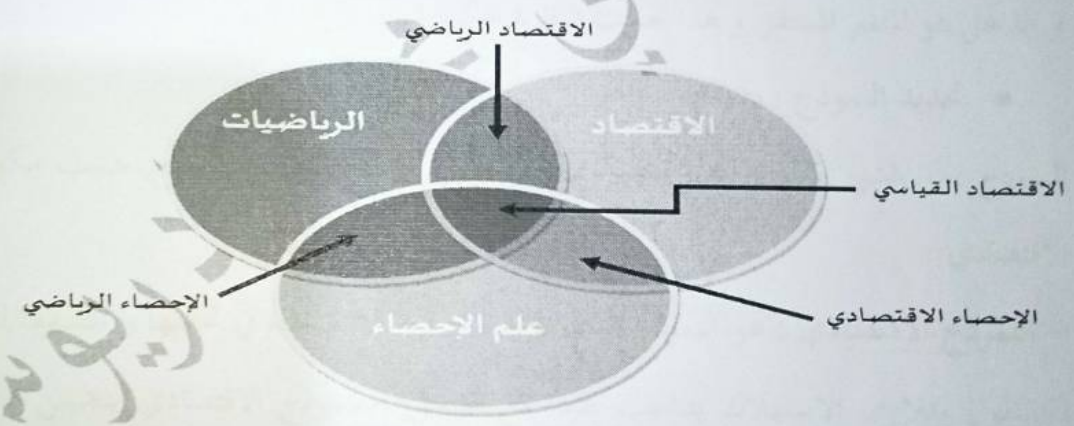
ومن هنا يمكن تعريف القياس الاقتصادي على انه التطبيق الكمي القياسي للنماذج الاحصائية و الرياضية باستخدام البيانات التاريخية لفحص النظريات الاقتصادية واختبار الفرضيات المتعلقة بالواقع الاقتصادي ،ثم استخدام هذه النماذج لغرض التنبؤ لمستبقل هذه البيانات و الاقتصاد ككل.

Scanned with Fast Scan

كما عرف كل من **ساموبلسون ،وكوبمان وستون** سنة 1954 على ان القياس الاقتصادي هو التحليل الكمي للظواهر الاقتصادية الفعلية ،وهذا التحليل مبني على النمو المتزامن للنظرية و المشاهدة المرتبط بالطرق الملائمة للاستدلال .

كما تم تعريفه من قبل **موريس الايس** على انه فرع من الاقتصاد يهدف الى دراسة الوسائل الاقتصادية على المستويين النظرية و التطبيقية بنفس المنطق المطبق على الفيزياء باستعمال نفس الطرق الكمية سواء كان على المستوى النظري او المستوى التطبيقي.

ومن خلال ماسبق يمكن ان نقول ان القياس الاقتصادي هو عبارة عن مزيج بين ثلاث عناصر ضرورية هي : البيانات الاقتصادية ،الطرق الرياضية و الاساليب الاحصائية ،لكن هذه العناصر ليست كافية لفهم العلاقات الكمية الحقيقية في الحياة الاقتصادية الحديثة لذلك يتطلب الامر مزج و اذابة كل من الاقتصاد و الرياضيات و الاحصاء في قالب متماسك يدعى القياس الاقتصادي. و يمكن ان نلخص مفهوم القياس الاقتصادي من خلال الشكل المولي :



ثالثا :اهداف الاقتصاد القياسي : يهدف القياس الاقتصادي الي تحقيق مجموعة من الاهداف هي:

1- اختبار العلاقات الاقتصادية بين المتغيرات كما تقدمها النظرية الاقتصادية وذلك من خلال الحصول على دليل علمي لاختبار القدرة التفسيرية لهذه العلاقات.

2- وضع و رسم السياسات الاقتصادية المناسبة :من خلال الحصول على التقديرات لمعالم العلاقات الاقتصادية والتي يمكن استخدامها فيما بعد في اتخاذ القرارات المناسبة ،وضع ورسم نهج للسياسات الاقتصادية المثلى. كل هذا سوف يساعد الحكومات و كذا اصحاب المؤسسات في اتخاذ القرارات المناسبة ،فمثلا الحكومة لا يمكن أن تتخذ قرارا بفرض ضريبة على سلعة معينة أو رفع الضريبة إلا بعد التعرف على مرونة العرض والطلب.

Scanned with Fast Scan

3- التنبؤ و توقع قيم المتغيرات قيد الدراسة من خلال استخدام التقديرات المتحصل عليها لمعالم العلاقات الاقتصادية في التنبؤ بسلوك القيم المستقبلية للمتغيرات ،هذا التنبؤ يساعد متخذ القرار فيما إذ كان من الضروري اتخاذ اجراءات معينة تتعلق بالمتغيرات الاقتصادية.

رابعا : منهجية البحث في الاقتصاد القياسي: تمر منهجية البحث في الاقتصاد القياسي بعدد من المراحل نذكر منها :

1- مرحلة صياغة النموذج و تحديد المتغيرات : و ذلك انطلاقا من طبيعة العلاقات الاقتصادية المختلفة ، كما ان تحديد المتغيرات يقصد بها المتغيرات المستعملة في النموذج وتنقسم الي :

- متغيرات تابعة او متغيرات مشروحة
- متغيرات مستقلة او متغيرات شارحة.

فعلى سبيل المثال في دالة الاستهلاك فان الدخل يعتبر هو المتغير الداخل في تحديد مستويات الاستهلاك عند الافراد و المجتمعات ،لذا يعتبر الانفاق الاستهلاكي دالة تابعة للدخل ومنه يكون الاستهلاك هو المتغير التابع و الدخل هو المتغير المستقل و هذا حسب النظرية االكينزية.

- **تحديد النموذج** :

النموذج ،هو الشئ المبسط لشئ معقد ويمكن ان نميز ثلاث انواع من النماذج حسب مكونات علم القياس الاقتصادي

- **النموذج الاقتصادي** : هو النموذج الذي تقترحه النظرية الاقتصادية في تحليلها و تفسيرها للظواهر الاقتصادية ،كاننقول مثلا ان الاستهلاك يتناسب طردا مع الدخل ،فالنموذج الاقتصادي يتضمن المتغيرات الاقتصادية بالاضافة الي نوع العلاقة (طردية او عكسية).

- **النموذج الرياضي** : هو عبارة عن تمثيل بسيط للنموذج الاقتصادي بدون استعمال التعابير اللفظية المعقدة و البيانات الاحصائية بل يقتصر على المعادلات الرياضية مثال في حالة الاستهلاك يمكن كتابة النموذج الرياضي المعبر عن العلاقة الطردية بين الاستهلاك و الدخل كمايلي : $0<B<1$, $C=a+By$ ،حيث C :هو الاستهلاك ،و y :هو الدخل.

- **النموذج القياسي** : هو عبارة عن نموذج رياضي و قياسي في نفس الوقت ،يستعمل بيانات احصائية واقعية تتضمن متغير عشوائي ،يفترض فيه ان يحقق مجموعة من الشروط ،مثلا في دالة الاستهلاك يمكن كتابة النموذج كمايلي: $0<B<1$, $C_i=a+By_i+U_i$ ، حيث C_i و y_i هو الاستهلاك و الدخل للعائلة i.

Scanned with Fast Scan

ونسمى كل من a,B معالم النموذج ،بينما تعبر U_i عن المتغير العشوائي او الخطأ و الذي يمكن ان تكون موجود في النموذج.

والفرق بين النموذج الرياضي و النموذج القياسي هو المتغير العشوائي ،حيث تكون العلاقة بين المتغيرات الاقتصادية في النموذج القياسي علاقة احتمالية غير مؤكدة ،كما ان النموذج القياسي يتميز كونه نموذج يسمح بتقدير العلاقة بين المتغيرات على عكس النموذج الرياضي و الاقتصادي اللذان يكتفيان بتحديد اتجاه العلاقة .

2- **مرحلة تقدير النموذج** :حيث تتضمن هذه المرحلة الخطوات الاتية :

- **تجميع البيانات** : وهي مجموعة من البيانات الاحصائية يتم تجميعها وفقا لشروط معينة ويمكن ان نميز بين عدة انواع من المتغيرات بحسب البيانات :

- **بيانات زمنية او سلاسل زمنية (time series data)** : وهي عبارة عن بيانات تتغير تبعا لمؤشر الزمن مثل الاستهلاك خلال الفترة 2000-2020 ،ويرمز لها بالرمز C_t .

- **بيانات القطعية او مقطعية (cross-sectional data)** : وهي عبارة عن بيانات و احصائيات تتغير وفقا لمؤشرات غير الزمن مثل تغير الاستهلاك حسب العائلات او حسب الدول حسب مداخيل الافراد او الشهادات ...الخ ،وذلك خلال فترة زمنية معلومة ثابتة ،ويرمز لها بالرمز C_i .

- **بيانات البانل (panal data)** : وهي عبارة عن بيانات احصائية متقاطعة ،حيث تتغير تبعا لمؤشر الزمن من جهة و لمؤشر غير الزمن من جهة اخرى ،مثل تغير الانفاق الاستهلاكي لمجموعة من الدول خلال مجموعة من السنوات ويرمز لها بالرمز C_{it} .

ومن بين الشروط التي يجب مراعاتها اثناء تجميع البيانات يمكن ذكر مايلي :

- ضرورة مرعاة الوحدة الزمنية :(بيانات سنوية ،نصف سنوية ،فصلية ،شهرية ...الخ) حسب نوع البيانات.
- مرعاة الوحدة المكانية : اي لكل منطقة جغرافية بيانات خاصة بها .
- تجانس البيانات: مثل نفس الوحدة او تحويل الوحدات الفيزيائية الى بيانات نقدية .
- ضرورة التفريق بين المتغيرات الحقيقية والمتغيرات الاسمية فالمتغيرات الحقيقية تلغي اثر السعر.
- ان تكون البيانات صادرة من مصدر موثوق.
- ضرورة فهم العلاقات المحاسبية سواء كان المستوى الجزئي و المستوى الكلي.

- **تقدير معالم النموذج** : نعني بتقدير معالم النموذج ايجاد مقدرات عددية ،وهناك عدة طرق منها : طريقة المربعات الصغرى ،طريقة المعقولية العظمى ،التقدير باستعمال مجال الثقة ،التقدير بطريقة العزوم...

Scanned with Fast Scan

3- مرحلة تقييم النموذج : وتنقسم هذه المرحلة الي :

- **التقييم الاقتصادي** : اي معرفة مدى ملائمة النتائج و قيم المعالم المقدرة للنظرية الاقتصادية ،فإذا كان هناك عدم تطابق مع النظرية الاقتصادية فهذا النموذج مرفوض اقتصاديا إلا اذا كان هناك تبرير لهذا الاختلاف كان تكون ظروف خاصة حالت دون تحقيق القاعدة النظرية فمثلا في حالة دالة الاستهلاك $C_i=a+By_i+U_i$ ،مع $0<B<1$ ،حيث B يمثل الميل الحدي الاستهلاك و الذي هو جزء من الدخل ،$a \geq 0$: حيث a :يمثل الاستهلاك التلقائي و هي قيمة الاستهلاك حتى و لو كان الدخل معدوما.

- **التقييم الاحصائي**: يقصد بالتقييم الاحصائي مدى الثقة الاحصائية للتقديرات الخاصة بمعالم النموذج ،واختبارات معنوية المعالم من الناحية الاحصائية و المعنوية الكلية للنموذج (عن طريق مقارنة الاحصائيات المحسوبة مع القيم الجدولية الموافقة لها) فاذا كانت المعالم معنوية احصائيا كذا النموذج فان النموذج مقبولا احصائيا و إلا فهو لا.

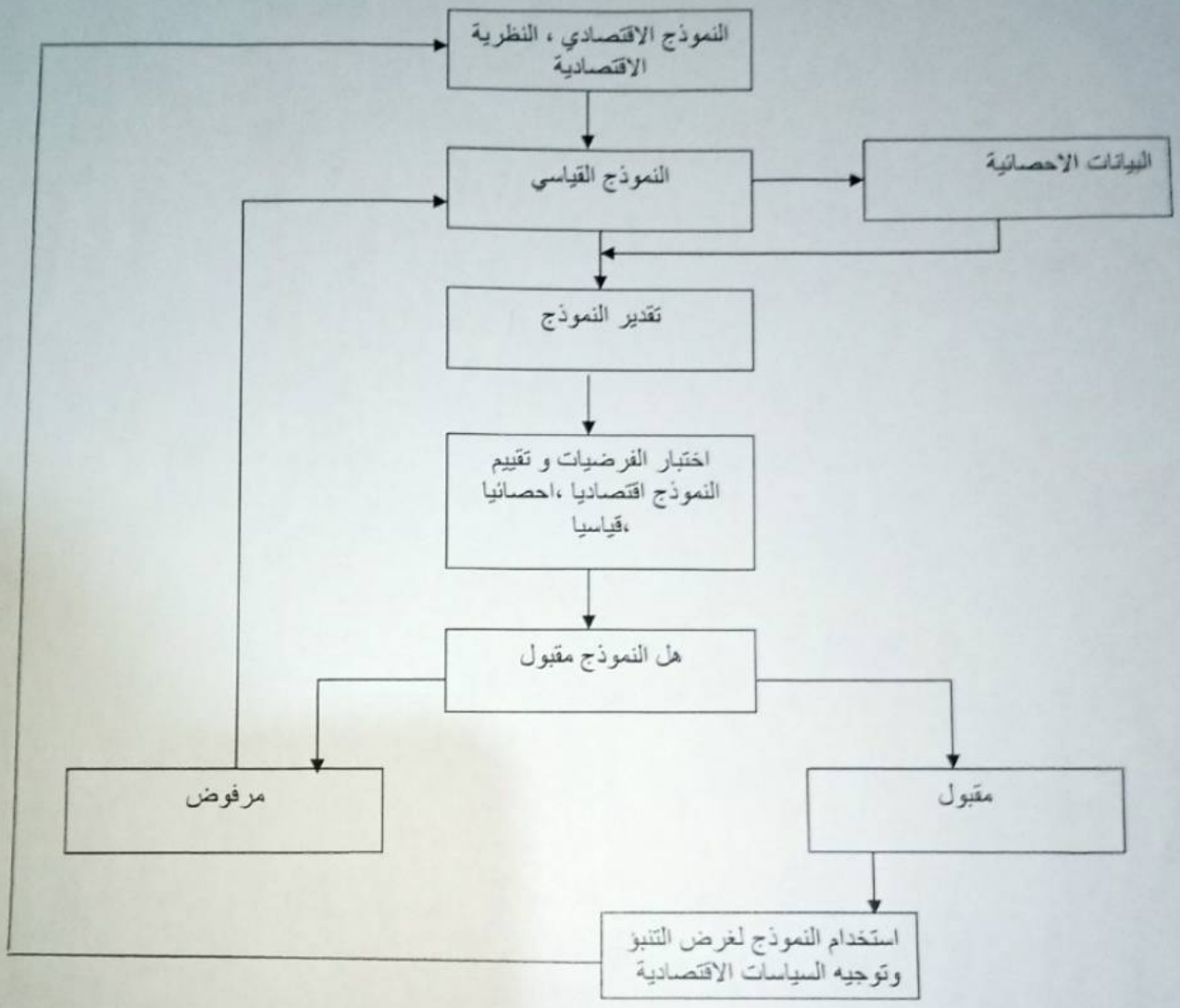
التقييم القياسي :يهدف التقييم القياسي الى اختبار مدى صحة الفرضيات التي تم بناء النموذج القياسي و المتعلقة خصوصا بالمتغير العشوائي ،فاذا كانت هذه الفرضيات محققة فان نتائج التقدير لها مصدقية كبيرة ويجعلها تتميز بصفات اساسية مهمة مثل ميزة عدم الانحياز و صفة الاتساق وفرضية عدم الارتباط بين الاخطاء و تجانس التباين.

4- مرحلة التنبؤ بالنموذج :يستخدم النموذج من اجل التنبؤ الذي يقوم على اساس المستقبل القريب و هو امتداد للماضي القريب ،ولكن اذا حدثت تغيرات هيكلية سريعة للظروف الاقتصادية و الاجتماعية فان النموذج لا يمكن ان يكون قادرا على التنبؤ بهذه المتغيرات.

من اجل معرفة مقدرة النموذج على التنبؤ لابد من اختبار استقرارية المعلمات المقدرة عبر الزمن و اختبار مدى حساسية هذه التقديرات لتغير حجم العينة المختارة.

ويمكن اختصار مراحل البحث في الاقتصاد القياسي بالمخطط الموالي :

Scanned with Fast Scan



Scanned with Fast Scan

و يسمى هذا النوع من الانحدار بالتقدير الخطي لان منحى الانحدار يكون دائما على شكل مستقيم (اي انتشار قيم المشاهدات المتغيرات يشكل سحابة مستقيمة).

وبشكل عام يمكن التعبير عن عن نموذج الانحدار الخطي البسيط قياسيا كما يلي :

$$Y_i = a + Bx_i + \varepsilon_i$$ ،حيث :

i: يمثل المشاهدة و التي تتغير من n.....1.2.3

Y_i : يمثل قيم المتغير التابع من n.....1.2.3

x_i : يمثل قيم المتغير المستقل من n.....1.2.3

a: تمثل الحد الثابت او الجزء المقطوع من المحور الرئيسي و هو عبارة عن قيمة المتغير التابع لما تنعدم قيمة المتغير المستقل

B: يمثل ميل خط المستقيم او معامل الانحدار و يعبر عن مقدار تغير في المتغير التابع اذا حدث تغير في المتغير المستقل بوحدوة واحدة ،وتبين علاقتها نوع العلاقة (طردية او عكسية)بين المتغير التابع و المستقل.

ε_i : يعبر عن الخطأ او المتغير العشوائي و الذي يتغير من n.....1.2.3 ،وهو الخطأ الذي يدخل في تفسير المتغير التابع ،وعليه يمكن كتابة انطلاقا من معادلة الانحدار الخطي البسيط العلاقة التالية :

$$\varepsilon_i = Y_i - (a + Bx_i)$$

ويرجع وجود الخطأ العشوائي الي :

- وجود اخطأ التحديد : مثا اهمال متغيرات مستقلة مهمة لها القدرة على تفسير المتغير التابع بشكل جيد او ادراج متغيرات مستقلة غير مهمة.

- وجود اخطأ القياس : كحدوث اخطاء في جمع البيانات و قياس المتغيرات الاقتصادية من جهة ،ومن جهة اخرى حدوث خطأ في صياغة النموذج كاعتبار ان العلاقة خطية بين المتغيرات و هي غير خطية في حقيقة الامر.

- وجود خطأ تذبذب العينة : يحدث هذا عند عدم تجانس بيانات و افراد العينة قيد الدراسة.

ثانيا : فرضيات النموذج :

من اجل بناء و تقدير نموذج الانحدار الخطي البسيط لبد من وضع مجموعة من الفرضيات الواجب اخذه بعين الاعتبار :

- **الفرضية 01** : النموذج خطي بالنسبة لمعالم و شعاع الاخطاء اي ان تمثيل معادلة الانحدار السابقة ياخذ شكل خط مستقيم .

- **الفرضية 02** : قيم المتغير المستقل حقيقية و ليست عشوائية و غير متساوية فيما بينها.

Scanned with Fast Scan

المحاضرة الثانية : نموذج الانحدار الخطي البسيط

اولا : تعريف نموذج الانحدار الخطي البسيط

يعتبر نموذج الانحدار الخطي البسيط ابسط انواع النماذج القياسية نظرا لاعتماده على متغير مستقل واحد لتفسير المتغير التابع ،و يمكن ان نذكر ان هذا النموذج لا يعبر في الكثير من الاحيان عن واقع السلوك المتغيرات و الظاهر الاقتصادية .لكن لايقلل من اهميتها في تفسير الكثير من العلاقات ،زيادة على انه يشكل منطلقا اساسيا نحو نماذج اخرى التي تم بنائها عليه من خلال اضافة متغيرات اخرى وتشكيل معادلات غير خطية.

وجاء كلمة انحدار Regression اول مرة من قبل الاقتصادي الانجليزي سير فرنسيس جالتون في اطار دراسته العلاقة بين طول الابناء وطول ابائهم حيث لاحظ ان طول الابناء يميل الى المعدل من خلال نموذج الانحدار الخطي بسيط ،والذي يمكن صياغته على النحو الاتي :

$$Child_i=a+Bfath_i$$

حيث يمثل $Child_i$ طول الابن باعتباره متغير تابع لأنه صفة وراثية من الاباء و ليس العكس ، $fath_i$يمثل طول الاباء باعتباره متغير مستقل يؤثر على طول الابناء وراثيا ،a يمثل الحد الثابت لطول الابناء ،B يمثل الميل الحدي لطول الابناء ،كما تمثل i عدد الافراد التي اجريت عليهم الدراسة ،ومن هنا يمكن تمثيل هذه العلاقة بالمعادلة الرياضية الاتية : $y=a+Bx$. ε

لكن الواقع الاقتصادي و الاجتماعي او الانساني كما في مثال جالتون لا يمتاز و لا يتشابه كما الرياضيات بالدقة و العلاقة كاملة التحقق ، حيث ان طول الاباء قد لا يؤثر على طول الابناء في حالة اصابة الابن بطفرة وراثية ،وبالتالي في هذه الحالة سنكون امام قيمة طول ابن (y) ليس هي تلك القيمة التي يتم حسابها وفق النموذج او المعادلة اعلاه ،اي ستكون قيمة y الحقيقية مختلفة الى حد ما عن قيمة y المحسوبة وفق المعادلة الرياضية ، وحتى تتمكن المعادلة الرياضية اعلاه من تمثيل الواقع نضيف اليها **حد اضافي** يهتم بالاختلالات بين القيم الحقيقية و القيم النظرية (المحسوبة) للمتغير التابع قيد الدراسة ،ويتمثل هذا الحد الاضافي في المتغير العشوائي او مايسمى بحد الخطأ والذي يرمز اليه عادة بالرمز ε_i ، فهذا الخطأ العشوائي يمثل في حقيقة الامر الفرق بين القيمة النظرية (المحسوبة) و القيمة الحقيقية للمتغير التابع ، ومن اجل الوصول الي افضل تقدير للعلاقة بين المتغير التابع و ذلك المستقل فيجب ان نجعل هذا الفرق بين القيمة النظرية و القيمة الحقيقية للمتغير التابع في ادنى (اقل) مستوى ممكن. و يمكن كتابة المعادلة النهائية للنموذج كمايلي : $\mathbf{Child_i=a+Bfath+\varepsilon_i}$

Scanned with Fast Scan

- **الفرضية 03** : كل قيم المتغير العشوائي في الفترة الزمنية تكون على الصدفة ،وقد تاخذ قيم (-) ،(+) ،(0).

- **الفرضية 04** : التوقع الرياضي للخطاء العشوائية معدوم $E(\varepsilon_i)=0$. اي ان الاخطاء لا تدخل في تفسير y وهذا يعني ايضا ان مجموع قيم ε_i يساوي الصفر . فاذا كان التوقع الرياضي للخطأ غير معدوم فهذا دليل على دخول هذا الخطأ في تفسير المتغير التابع مما يفقد اثر المتغير المستقل على المتغير التابع

- **الفرضية 05** : تجانس (ثبات) تباين الاخطاء ،اي ان تشتت الاخطاء حول المتوسط الحسابي لها ثابت طيلة فترة الدراسة اي ان قيم ε_i تتغير في مدى ثابت او في حدود ثابتة حول متوسطها الحسابي (0) وهذا يعني ان المدى بين الحد الاقصى و الحد الادنى لقيم الخطأ العشوائي عند كل قيمة للمتغير x_i ثابت اي بعبارة اخرى $E(\varepsilon_i^2)=\delta_\varepsilon^2$.

- **الفرضية 06** : الاخطاء مستقلة عن بعضها البعض اي عدم وجود ارتباط بين الاخطاء بمعنى قيمة الخطأ(ε_i) التي باخذها في فترة ما لا تكون مرتبطة او متعلقة بقيمة خطأ ما في فترة سابقة او فترة لاحقة ، وذلك من اجل ان يكون x_i وحده المتغير الذي يؤثر في المتغير التابع y اي بعبارة اخرى $cov(\varepsilon_i, \varepsilon_j)=0$ و نكتب :

$E(\varepsilon_i-E\varepsilon_i)(\varepsilon_j-E\varepsilon_j)=E(\varepsilon_i\varepsilon_j)=0$ ،حيث i و j تمثلان فترتين زمنيتين او فردين مختلفين .

- **الفرضية 07** : الاخطاء غير مرتبطة بالمتغير المستقل $cov(x_i, \varepsilon_i)=0$ ،لان ارتباط المتغير المستقل بالخطأ فهذا يعني ان المتغير المستقل يدخل في تفسير الخطأ مما يجعل هذا الخطأ لا يعبر فقط عن ابتعاد القيمة المحسوبة عن القيمة الحقيقية للمتغير التابع.

- **الفرضية 08** : المتغير العشوائي يتبع التوزيع الطبيعي بمتوسط حسابي يساوي الصفر و تباين ثابت لا يتغير عبر الزمن اي :$\varepsilon_i \sim N(0, \delta_\varepsilon^2)$.

- **الفرضية 09**: عدد المشاهدات في الانحدار الخطي البسيط اكبر من 2 اي $n>2$ ،بينما في الانحدار المتعدد تكون $n>k+1$ ،حيث k يمثل عدد المتغيرات المستقلة ،حيث انه كل ما كان عدد المشاهدات كبير كلما اتجهت الاخطاء نحو الصفر و بالتالي يصبح المتغير المستقبل له اكبر تعبير للمتغير التابع .

ثالثا : تقدير معالم النموذج :

- اذا كان لدينا النموذج الانحدار الخطي البسيط الاتي :

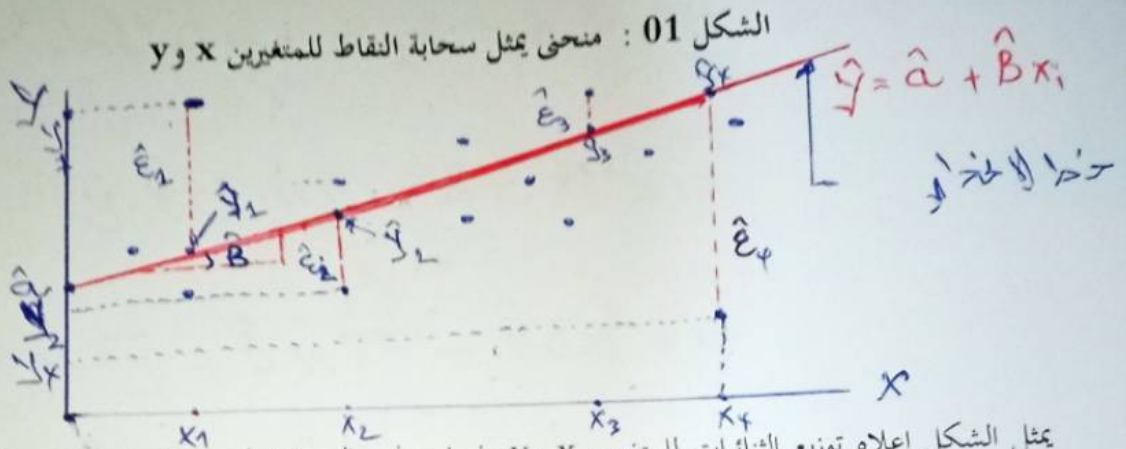
$$y_i=a+Bx_i+\varepsilon_i$$ ،حيث :

i: يمثل المشاهدة وتتغير من 1.2.3.....n ، y_i: يمثل قيم المتغير التابع ،x_i : يمثل قيم المتغير المستقل ،a و B تمثل *معلمات النموذج المراد تقديرهما* ،ε_i تمثل الخطأ العشوائي.

Scanned with Fast Scan

ويقصد بتقدير النموذج ايجاد **مقدارات** المعالم a و B بدلالات x و y و يرمز لها بالرمز $\hat{a}$ و $\hat{B}$ على الترتيب و هي احسن القيم الممكنة للتعبير عن النموذج في الصيغة الخطية و النموذج المكتوب بدلالة قيم $\hat{a}$ و $\hat{B}$ يسمى النموذج المقدر ، ويرمز للقيمة المقدرة لـ y_i(القيمة الحقيقية) بالرمز $\hat{y}_i$.

الشكل 01 : منحنى يمثل سحابة النقاط للمتغيرين x و y



يمثل الشكل اعلاه توزيع الثنائيات للمتغيرين x و y ،فيما يعرف بالسحابة التي تتمثل في المساحة التي تحددها النقاط في الشكل ،حيث تهدف عملية التقدير الى ايجاد امثل خط انحدار الذي يمر بهاته السحابة مقلصا الاخطاء الي اقصى حد ممكن ،حيث من غير الممكن ايجاد خط انحدار يمر بجميع النقاط تكون عنده الاخطاء معدومة ،و عليه نقوم بإيجاد افضل خط (الخط الاحمر في الشكل) تكون عنده الاخطاء عند ادنى مستوى او معدومة ،وكما هو مبين في الشكل اعلاه فان الفجوة بين المشاهدة الحقيقية (النقطة) و اسقاطها على خط الانحدار هو يعبر عن الخطأ ، وعليه يمكن تسمية المشاهدات (النقاط) ب y اي القيمة الحقيقية ونسمي اسقاطها على خط الانحدار ب $\hat{y}$ اي القيمة النظرية او المحسوبة .و بالتالي يكون($\varepsilon_i = y_i - \hat{y}_i$).

– طريقة المربعات الصغرى :

تعتبر طريقة المربعات الصغرى من بين افضل الطرق المستخدمة في نماذج الانحدار،فهي تحاول ايجاد احسن تصحيح خطي عن طريق تدنية الفوارق بين القيم الحقيقية و القيم النظرية للمشاهدات التي يمنحها مستقيم خط الانحدار للنموذج المقدر،غير ان هذه الفوارق تأخذ احيانا الاشارة السالبة و احيانا الاشارة الموجبة ،وبطريقة رياضية نكتب $\text{Min } \Sigma (\varepsilon_i) = \text{Min } \Sigma (y_i - \hat{y}_i)$.

ولتجاوز مشكل الاشارة نلجأ الى استعمال مربعات الفوارق او ما يعرف مربع الانحرافات ،مما يعطينا مجموعة انحرافات موجبة ،ومن ثم نعمل على تقليل مجموع مربعات الانحرافات قدر الامكان اي :

$$\text{Min } \Sigma(\varepsilon_i)^2 = \text{Min } \Sigma (y_i - \hat{y}_i)^2 = \text{Min } \Sigma(yi - \hat{a} - \hat{B}\, xi)^2$$

Scanned with Fast Scan

.. حتى نتمكن من تصغير المعادلة S. الى ادنى قيمة او مستوى
فاننا علينا سوف البحث عن المشتقة الجزئية للمعادلة S.
بالنسبة للمجاهيل a، B مع ، ونضع المشتقة تساوي الصفر
والحل على النحو التالي:

$$\begin{cases} \frac{dS}{da} = \frac{d\sum_{i=1}^{n}(y_i - a - Bx_i)^2}{da} = -2\sum_{i=1}^{n}(y_i - \hat{a} - \hat{B}x_i) = 0 \\ \frac{dS}{dB} = \frac{d\sum_{i=1}^{n}(y_i - a - Bx_i)^2}{dB} = -2\sum_{i=1}^{n} x_i(y_i - \hat{a} - \hat{B}x_i) = 0 \end{cases}$$

لأن مشتقة الدالة $(F^n)' = nF'F^{n-1}$ حيث تمثل F
الدالة المراد اشتقاقها، في هذه الحالة تمثل الدالة S.
و n هو اس الدالة، و (2) في هذه الحالة، ومن اجل
ايجاد قيم $\hat{a}$، $\hat{B}$، كمتغير الدالة نحذف اثنان مع نشر
المجموع $[\Sigma]$ على الاقواس مع العلم ان $\Sigma a = na$
ومنه نحصل على المعادلتين الآتيتين:

$$\begin{cases} \Sigma y_i - n\hat{a} - \hat{B}\cdot\Sigma x_i = 0 \quad \cdots\cdots (1) \\ \Sigma x_i y_i - \hat{a}\Sigma x_i - \hat{B}\Sigma x_i^2 = 0 \quad \cdots (2) \end{cases}$$

ما يجب ملاحظته انه تم تغيير كتابة كل من B، a الى
$\hat{B}$، $\hat{a}$ على التوالي، وهذا للدلالة على انها قيمتين
محسوبتين. مثل القيم المحسوبة لـ y
نعلم ان: $\bar{x} = \Sigma x_i / n$ ومنه $\Sigma x_i = n\bar{x}$ - وفس الشيء
بالنسبة لـ y
بالتعويض في المعادلتين نجد:

المعادلة ①:

$$\sum y_i - n\hat{a} - \hat{\beta}\sum x_i = 0 \Rightarrow n\bar{y} = n\hat{a} + n\hat{\beta}\bar{x}$$

$$\Rightarrow \bar{y} = \hat{a} + \hat{\beta}\bar{x}$$

$$\Rightarrow \boxed{\hat{a} = \bar{y} - \hat{\beta}\bar{x}}$$

المعادلة ②: نقوم بتعويض قيمة $\hat{a}$ في المعادلة ② فنجد:

$$\sum x_i y_i - (\bar{y} - \hat{\beta}\bar{x}) \cdot n\bar{x} - \hat{\beta}\sum x_i^2 = 0$$

$$\sum x_i y_i - n\bar{x}\bar{y} + \hat{\beta} n\bar{x}^2 - \hat{\beta}\sum x_i^2 = 0$$

$$\sum x_i y_i - n\bar{x}\bar{y} - \hat{\beta}\left[\sum x_i^2 - n\bar{x}^2\right] = 0$$

$$\sum x_i y_i - n\bar{x}\bar{y} = \hat{\beta}\left[\sum x_i^2 - n\bar{x}^2\right] = 0 \Rightarrow \boxed{\hat{\beta} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}} \quad ②'$$

$$\Rightarrow \boxed{\hat{\beta} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}} \quad ②''$$

لنعوض ②' على n نجد:

$$\Rightarrow \boxed{\hat{\beta} = \frac{cov(x_i, y_i)}{V(x_i)}} \quad ②'''$$

وتسمى طريقة حساب $\hat{\beta}$ وفق المعادلة ②'' طريقة العزوم

كما يمكن حساب $\hat{\beta}$ أيضا بالعلاقة التالية ②'' طريقة التوفيق حول المحاور

$$\boxed{\hat{\beta} = \frac{n\sum x_i y_i - \sum x_i \cdot \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}} \quad ②''''$$

ملاحظة: في حالة عدم وجود الحد الثابت $\hat{a}$ نحسب حساب $\hat{\beta}$ كما يلي:

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

Scanned with Fast Scan

7

مثال 1: لدينا x_i و y_i متغيرتان بينهما علاقة خطية. المطلوب:
1. تعيين معادلة خط الانحدار بطريقة المربعات الصغرى.
2. تقدير قيمة y عندما $x = 5$.

x_i	y_i	$(x_i - \bar{x})^2$	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$\hat{y}_i$
1	4	9	-3	-3	9	2,929
4	5	0	0	-2	0	4,000
5	7	1	1	0	0	8,357
6	12	4	2	5	10	9,714
Σ=16	28	Σ=14	0	0	19	

- قدر معادلة خط الانحدار لهذه العينة "أي أوجد $\hat{a}$، $\hat{B}$ بطريقة المربعات الصغرى".
- احسب قيمة $\hat{y}$ لكل من $\hat{a}$، $\hat{B}$.
- احسب القيم النظرية للمتغير التابع.

الحل:

1. حساب المتوسط الحسابي لكل من x_i و y_i:

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{16}{4} \Rightarrow \bar{x} = 4$$

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{28}{4} = 7$$

2. قيمة ميل المستقيم بطريقة المربعات الصغرى:

$$\hat{B} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2}$$

$$\hat{B} = \frac{19}{14} = 1,357$$

$$\hat{a} = \bar{y} - \hat{B}\bar{x} \Rightarrow \hat{a} = 7 - 1,357 \times 4 \Rightarrow \hat{a} = 1,572$$

وعليه معادلة خط الانحدار تكتب كما يلي: $\hat{y} = 1,572 + 1,357 x_i$

Scanned with Fast Scan

8

ب. حساب قيم $\hat{y}_i$ (القيمة التقديرية) للعينة الحقيقية y_i:

$\hat{y}_1 = 1,572 + 1,357(1) = 2,929$

$\hat{y}_2 = 1,572 + 1,357(4) = 7,000$

$\hat{y}_3 = 1,572 + 1,357(5) = 8,357$

$\hat{y}_4 = 1,572 + 1,357(6) = 9,714$

ج. اعط قراءة لكل من $\hat{\alpha}$، $\hat{\beta}$:

$\hat{\alpha}$: يمثل الحد الثابت (القاطع)، وهو القيمة التي نحصل عليها y لما $x = 0$

$\hat{\beta}$: يمثل الميل الحدي لمعادلة خط الانحدار، وهو يمثل مقدار التغير الذي يمكن أن نحصل في y إذا تغير x بوحدة واحدة، مع بقاء باقي المتغيرات ثابتة.

المحاضرة الثالثة: نموذج الانحدار الخطي البسيط (تابع)

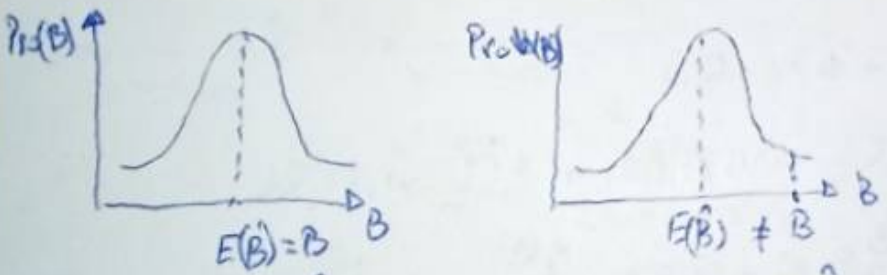
أولا: الخواص الإحصائية لمقدرات MCO: $\hat{a}$ ، $\hat{B}$

1- خاصية عدم التحيز:

تمتلك خاصية عدم التحيز في ضرورة ان يكون التوقع الرياضي للمقدرة يساوي تماما المعلمة الحقيقية اي بعبارة اخرى: $E(\hat{B}) = B$

اي الفرق بينهما: $E(\hat{B}) - B = 0$ ، حتى نفهم هذه الخاصية ندرج الشكل الآتي:

حالة الغالبية (عدم الاتساق) — عدم التحيز $E(\hat{B}) = B$ — تحيز $E(\hat{B}) \neq B$



$\hat{B}$ متحيز لـ B — $\hat{B}$ غير متحيز لـ B

وهذه الخاصية تنطبق ايضا على الحد الثابت اي:

$E(\hat{a}) = E(a) = a$

البرهان: لدينا:

- بالنسبة: $\hat{B}$

$$\hat{B} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

نضع $w_i = \frac{(x_i - \bar{x})}{(x_i - \bar{x})^2}$

$$\sum w_i = \frac{\sum (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \quad (1)$$

$$\sum w_i (y_i - \bar{y}) = \frac{\sum (x_i - \bar{x}).(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \hat{B}$$

نضرب (1) في $(y_i - \bar{y})$ نجد:

$\hat{B} = \frac{\sum w_i (y_i - \bar{y})}{\sum w_i (x_i - \bar{x})} = \sum w_i y_i - \sum w_i \bar{y}$

وعليه:

$= \sum w_i (a + B x_i + \varepsilon_i)$

$= \sum w_i a + B \sum w_i x_i + \sum w_i \varepsilon_i$

$\hat{B} = B + \sum w_i \varepsilon_i$ من الفرضيات

$E(\hat{B}) = E(B) = B$

ومنه نكتب:

وعليه $\hat{B}$ غير متحيز لـ B

بالنسبة لـ $\hat{a}$ لدينا:

$\hat{a} = \bar{y} - \hat{B}\bar{x}$

$= \frac{\sum y_i}{n} - (\sum w_i y_i) \cdot \bar{x}$

$= \sum y_i \left(\frac{1}{n} - w_i \bar{x}\right)$

$= \sum \left(\frac{1}{n} - w_i \bar{x}\right) \cdot (a + B x_i + \varepsilon_i)$

$= \sum \left(\frac{1}{n} - w_i \bar{x}\right) a + \sum \left(\frac{1}{n} - w_i \bar{x}\right) B x_i + \sum \left(\frac{1}{n} - w_i \bar{x}\right) \varepsilon_i$

$= \frac{na}{n} - a \sum w_i \bar{x} + \frac{B}{n} \sum x_i - B\bar{x} \sum w_i x_i + \frac{\sum \varepsilon_i}{n} - \bar{x} \sum w_i \varepsilon_i$

$= a + B\bar{x} - B\bar{x} - \bar{x} \sum w_i \varepsilon_i$

$\hat{a} = a + \bar{x} \cdot \sum w_i \varepsilon_i$

وعليه نكتب:

$E(\hat{a}) = E(a) - \bar{x} \sum w_i \cdot E(\varepsilon_i)$ ومنه،

$E(\hat{a}) = E(a) = a$

وعليه $\hat{a}$ غير متحيز لـ a.

Scanned with Fast Scan

③

خصائص أفضل تباين (أقل تباين) لـ: α, β :

نعلم تباين كل من $\hat{\alpha}, \hat{\beta}$ على التوالي بالعلاقة التالية:

$$V(\hat{\beta}) = \frac{\sigma_{\varepsilon_i}^2}{\sum (x_i - \bar{x})^2}$$

$$V(\hat{\alpha}) = \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}\right] \cdot \sigma_{\varepsilon_i}^2$$

حتى يكون لـ α, β أفضل تباين يجب أن يكون:

$$\lim_{n \to \infty} V(\hat{\beta}) = 0$$

$$\lim_{n \to \infty} V(\hat{\alpha}) = 0$$

بالنسبة لـ $V(\hat{\beta})$

$$\lim_{n \to \infty} V(\hat{\beta}) = \lim_{n \to \infty} \left(\frac{\sigma_{\varepsilon_i}^2}{\sum (x_i - \bar{x})^2}\right)$$

$$= \lim_{n \to \infty} \frac{\sigma_{\varepsilon_i}^2 / n}{\sum (x_i - \bar{x})^2 / n}$$

من الفرضيات لدينا: $\sigma_{\varepsilon_i}^2 = A$ ثابت مهما تغير حجم العينة

$\lim_{n \to \infty} \sigma_{\varepsilon_i}^2 / n = 0$ ، عليه: - n - أي مهما يزداد

كما أن $\lim_{n \to \infty} \frac{\sum (x_i - \bar{x})^2}{n} \neq 0$ حيث أن قيم x_i مختلفة $\sum (x_i - \bar{x})^2 \neq 0$

فيما يتعلق بصب الفرضية

إذن يكون لدينا: $\lim_{n \to \infty} V(\hat{\beta}) = 0$

Scanned with Fast Scan

بالنسبة لـ $V(\hat{a})$: $\lim_{n\to\infty} \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\right] \sigma_\varepsilon^2$

$$= \lim_{n\to\infty} \left[\frac{\sigma_\varepsilon^2}{n} + \frac{\bar{x}^2 \cdot \sigma_\varepsilon^2}{\sum(x_i-\bar{x})^2}\right]$$

$$\lim_{n\to\infty} \frac{\sigma_\varepsilon^2}{n} = 0.$$

$$\lim_{n\to\infty} \frac{\bar{x}^2 \cdot \sigma_\varepsilon^2 / n}{\sum(x_i-\bar{x})^2/n} = 0 \Rightarrow \lim_{n\to\infty} \bar{x}\sigma_\varepsilon^2/n = 0$$

و

$$\lim_{n\to\infty} \sum(x_i-\bar{x})^2/n \neq 0$$

، طبعاً :

$$\lim_{n\to\infty} V(\hat{a}) = 0$$

3. خاصية 3 سماة "الفعالية"

نقول عن المقدرتين $\hat{a}$ ، $\hat{B}$ لمعاملة ، نتيجة متساقتين

إذا حققتا الشرطين :

$$E(\hat{B}) = B \; ; \; E(\hat{a}) = a$$

$$\lim_{n\to\infty} V(\hat{B}) = 0 \; ; \; \lim_{n\to\infty} V(\hat{a}) = 0$$

ملاحظة : لدينا :

$$V(\hat{a}) = \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\right]\sigma_\varepsilon^2 = \left[\frac{\sigma_\varepsilon^2}{n} + \frac{\bar{x}^2\sigma_\varepsilon^2}{\sum(x_i-\bar{x})^2}\right] \quad (1)$$

$$V(\hat{a}) = \left[\frac{\sigma_\varepsilon^2}{n} + \bar{x}^2 \cdot V(\hat{B})\right]$$

(2) : حساب كتابة التباين المشترك $(a, \hat{B})$ على الشكل التالي :

$$Cov(a,\hat{B}) = -\bar{x} \cdot \frac{\sigma_\varepsilon^2}{\sum(x_i-\bar{x})^2} = -\bar{x} \cdot V(\hat{B})$$

Scanned with Fast Scan

تباين الاخطاء

سبق ان تباين كل متغير $V(B)$ ، $V(\hat{B})$ غير معروف لانه يرتبط بتباين الخطأ العشوائي "المتغير" $\sigma^2_{\varepsilon_i}$. وعليه يجب تقدير تباين الاخطاء من اجل الحصول على تباين المعالم:

ويخص تباين الخطاء بالمعادلة الآتية:

$$E\hat{\varepsilon}_i^2 = \sigma^2_{\varepsilon_i} + (n-1)\sigma^2_{\varepsilon_i} - 2\sigma^2_{\varepsilon_i}$$

$$= (n-2)\sigma^2_{\varepsilon_i} \Rightarrow \hat{\sigma}^2_{\varepsilon_i} = \frac{E\hat{\varepsilon}_i^2}{n-2}$$

وهذه المعادلة غير متحيز، حيث $E(\hat{\varepsilon}_i^2)$: تعبر عن مجموع مربعات البواقي حيث: $E\hat{\varepsilon}_i^2 = (y - \hat{y}_i)^2$

$n-2$: درجة الحرية، تعبر عن حجم العينة $(n-2)$ و (2) يعني لوجود معلمتين للمعادلة في النموذج

$\sigma^2_{\varepsilon_i}$: هذا يقيس الخرف القيم الفعلية من القيم المقدرة وإذا كان هذا التقدير كبير فهذا يعني ان بعد او الخرف بين القيم الفعلية عن القيم المحسوبة الخاصة بالمتغير التابع هو كبير مما يجعل النموذج غير مثالي او غير صحيح

Scanned with Fast Scan

- حالة المعلمتين α و β:

1- في حالة $n \geq 30$. يمكن في هذه الحالة

بمعرفة توزيع $\hat{\alpha}$ و $\hat{\beta}$ يمكننا تحديد مجال الثقة و إجراء اختبار الفرضيات

الذي موضوعة حول α و β على التوالي . وتعطي مجالات القيم التي يمكن

أن تأخذها المعالم الحقيقية للمجتمع، حيث نضع عند كل مجال

مستوى ثقة الذي يكون $(1-\alpha)$ أي واحد مطروح منه مستوى

المعنوية (الخطر) .

وفي حالة عينات أقل من 30 تباين σ مجهول غير معروف

نجد المعلمتين α و β يكون تحديدها بإتباع توزيع ستودنت

حيث يعطي مجالهما كما سبق على التوالي وذلك عند مستوى معنوية α

$$a \in \left[\hat{a} - \hat{\sigma}_{\hat{a}} \cdot t_{n-2,\,1-\frac{\alpha}{2}} ; \hat{a} + \hat{\sigma}_{\hat{a}} \cdot t_{n-2,\,1-\frac{\alpha}{2}} \right]$$

$$B \in \left[\hat{B} - \hat{\sigma}_{\hat{B}} \cdot t_{n-2,\,1-\frac{\alpha}{2}} ; \hat{B} + \hat{\sigma}_{\hat{B}} \cdot t_{n-2,\,1-\frac{\alpha}{2}} \right]$$

2- في حالة $n < 30$: فإن المعلمتين α و β يمكننا تشكيل مجال

ثقتهما باستخدام توزيع z (التوزيع الطبيعي) حيث نمكن مجال

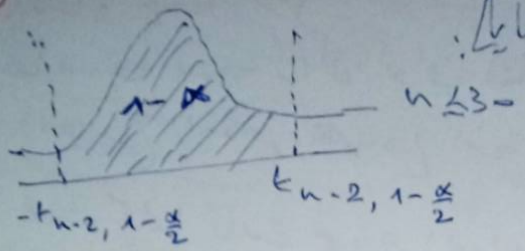
الثقة لكل المعلمتين على التوالي عند مستوى معنوية α

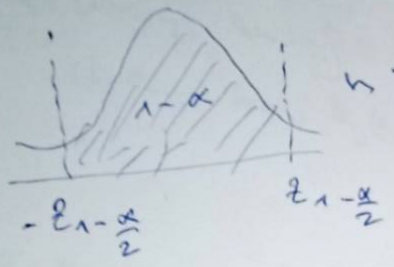
$$a \in \left[\hat{a} - \hat{\sigma}_{\hat{a}} \cdot z_{1-\frac{\alpha}{2}} ; \hat{a} + \hat{\sigma}_{\hat{a}} \cdot z_{1-\frac{\alpha}{2}} \right]$$

$$B \in \left[\hat{B} - \hat{\sigma}_{\hat{B}} \cdot z_{1-\frac{\alpha}{2}} ; \hat{B} + \hat{\sigma}_{\hat{B}} \cdot z_{1-\frac{\alpha}{2}} \right]$$

Scanned with Fast Scan

تمثيل مجال الثقة بيانيا كما يلي:





بالنسبة لـ σ_{ei}: مجال الثقة بالنسبة لتباين الخطأ (σ^2_{ei}):

يمكن اعطاء توزيع تباين الخطأ كما يلي:

$$\sigma^2_{ei} \in \left[\frac{(n-2).\hat{\sigma}^2_{ei}}{\chi^2_{n-2,\frac{\alpha}{2}}} \; ; \; \frac{(n-2).\hat{\sigma}^2_{ei}}{\chi^2_{n-2,1-\frac{\alpha}{2}}} \right]$$

حيث $n-2$: ق د ح

α: مستوى المعنوية

$\hat{\sigma}^2_{ei}$: ق تباين الخطأ المقدر

$\chi^2_{n-2,1-\frac{\alpha}{2}} - \chi^2_{n-2,\frac{\alpha}{2}}$: ق قيمتا اكاي مربع

التي يمكن استخراجها من الجدول (اكاي مربع) χ^2

Scanned with Fast Scan

النموذج المقدر للنموذج الحقيقي وكلما يسمى بجودة التوفيق

حيث كلما كانت جودة البواقي كبيرة قلت جودة التمثيل
والعكس صحيح. بمعنى ان البواقي تقدم عن المتغير التابع
والذي يحسن لها يبعد حول وسطه.

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$$

$$y_i = \hat{y}_i + \hat{\varepsilon}_i \quad -(1)$$

وعليه

$$(2) \Rightarrow y_i - \bar{y} = \hat{y}_i - \bar{y} + \hat{\varepsilon}_i$$

$$\Rightarrow (y_i - \bar{y})^2 = (\hat{y}_i - \bar{y})^2 + \hat{\varepsilon}_i^2$$

$$\Rightarrow \boxed{\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum \hat{\varepsilon}_i^2} \quad -(*)$$

حيث المعادلة (*) تستخدم لقياس القدرة التفسيرية للنموذج حيث:

$\sum (y_i - \bar{y})^2$ يسمى مجموع مربعات الانحرافات الكلية

Total Sum of Squares (T.S.S)

Regression Sum of Squares (R.SS) : يسمى مجموع مربعات الانحرافات

Error Sum of Squares (E.SS) : مجموع مربعات البواقي (الأخطاء)

Scanned with Fast Scan

إعادة صياغة المعادلة (*) على الشكل التالي:

$$TSS = RSS + ESS$$

وبقسمة طرفي المعادلة على TSS نجد:

$$\frac{TSS}{TSS} = \frac{RSS}{TSS} + \frac{ESS}{TSS} \Rightarrow 1 = \frac{RSS}{TSS} + \frac{ESS}{TSS}$$

$$\Rightarrow 1 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} + \frac{\sum \hat{\varepsilon}_i^2}{\sum(y_i - \bar{y})^2}$$

حيث تسمى النسبة $\frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$ بمعامل التحديد " Coefficient of determination "

ويرمز له بالرمز (R^2)

وعليه يمكن معامل التحديد R^2 من العلاقة السابقة

$$R^2 = 1 - \frac{\sum \hat{\varepsilon}_i^2}{\sum(y_i - \bar{y})^2}$$

ومعامل التحديد R^2 يقيس ويشرح نسبة تأثر التغيرات الحاصلة في المتغير التابع (y) لتغيرات المتغير المستقل (x). وعليه فإن R^2 جزئية تأثير المتغير المستقل في المتغير التابع، وبالتالي فهو يقيس القدرة التفسيرية للنموذج وهو محصور $0 \leq R^2 \leq 1$ وكلما كان قريب من الواحد R^2 يدل ذلك على جودة النموذج كما نأخذ والعكس صحيح، كما يمكن كتابة R^2 بالعلاقات التالية:

$$R^2 = \frac{\hat{\beta}^2 \cdot \sum(x_i - \bar{x})^2}{\sum(y_i - \bar{y})^2} = \frac{\hat{\beta}^2 [\sum x_i^2 - n\bar{x}^2]}{\sum y_i^2 - n\bar{y}^2} = \frac{\hat{\beta} \cdot \sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(y_i - \bar{y})^2}$$

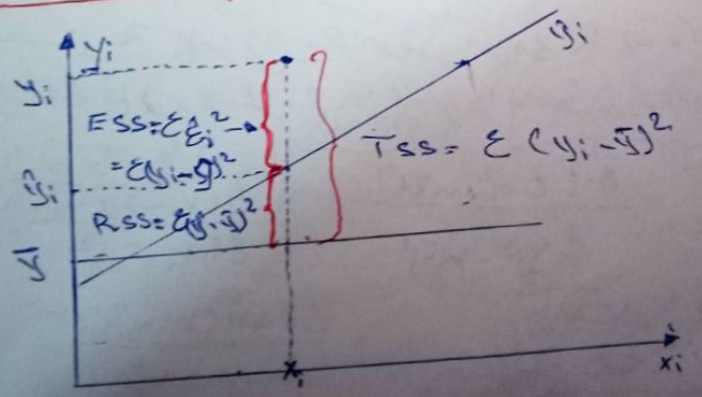
$$= \frac{[\sum(x_i - \bar{x}) \cdot (y_i - \bar{y})]^2}{\sum(x_i - \bar{x})^2 \cdot \sum(y_i - \bar{y})^2} = \frac{[Cov(x_i, y_i)]^2}{Var(x_i) \cdot Var(y_i)}$$

Scanned with Fast Scan

معامل ارتباط: هو مقياس يبين الخطر التربيعي لمعامل
$r = \sqrt{R^2}$ ، والفرق بين معامل التحديد ومعامل ارتباط (r) يكمن
في السببية، حيث يقيس معامل ارتباط العلاقة الموجودة بين المتغيرين
بغض النظر عن الدور الذي يلعبه كل متغير (متغير تابع أو مستقل)
وهو محصور بين $-1 \leq r \leq 1$ ، ويقيس قوة العلاقة بين المتغيرين
في حالة الانحدار الخطي البسيط يعطى بالعلاقة:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}} = \frac{Cov(x_i, y_i)}{\sqrt{Var(x_i) \cdot Var(y_i)}} = \sqrt{R^2}$$

مصدر التغير	مجموع المربعات	درجة الحرية	متوسط مجموع المربعات
المتغير المستقل الناتج	$RSS = \sum (\hat{y}_i - \bar{y})^2$	1	$\sum (\hat{y}_i - \bar{y})^2 / 1$
الخطأ	$ESS = \sum \hat{\varepsilon}_i^2$	$n-2$	$\sum \hat{\varepsilon}_i^2 / n-2$
المجموع	$TSS = \sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum \hat{\varepsilon}_i^2$	$n-1$	/



Scanned with Fast Scan

تقييم الاحصائي للنموذج

بقاء على مرحلة تقدير معالم النموذج نشرع في المرحلة الثانية من مراحل الدراسة القياسية وهي مرحلة اختبار جودة تمثيل النموذج، أو ما يعرف باختبار معنوية المعالم المقدرة الخاصة بالنموذج، ويقصد باختبار معنوية المعالم للنموذج هل هذه المعالم معنوية أو غير معنوية أي هل B و a كافية لتمثيل Y و X على التوالي

أهم الاختبارات:

1- اختبار ستودنت: يستخدم اختبار ستودنت للتحقق من معنوية المعالم المقدرة للنموذج القياسي الذي لدينا، انطلاقا من التوزيع الاحصائي لستودنت، حيث يعتمد هذا الاختبار على توفر شرطين هما:

بالنسبة (B)

$$\begin{cases} H_0: B = 0 \\ H_1: B \neq 0 \end{cases}$$

من أجل اختبار فرضية العدم H_0، نقوم بحساب الاحصائية $t_c = \frac{\hat{B} - B}{\sigma_{\hat{B}}} \sim t_{n-2}$، ومقارنتها بالاحصائية الجدولية المتوافقة معها في جدول ستودنت، وفق فرضية H_0 يكون لدينا: $t_c = \frac{\hat{B}}{\sigma_{\hat{B}}}$

- فإذا كانت $t_c = \frac{|\hat{B}|}{\sigma_{\hat{B}}} > t_{n-2, 1-\frac{\alpha}{2}}$ عند مستوى معنوية α٪ ودرجة حرية $(n-2)$

ففي هذه الحالة: H_0 مرفوض، H_1 مقبول، أي المعلمة B تختلف عن الصفر

- فإذا كانت $t_c = \frac{|\hat{B}|}{\sigma_{\hat{B}}} < t_{n-2, 1-\frac{\alpha}{2}}$ عند مستوى معنوية α٪ ودرجة الحرية $(n-2)$

ففي هذه الحالة: H_0 مقبول، H_1 مرفوض أي المعلمة B غير معنوية، بمعنى أنه نأخذ القيمة (0) أي أن خط الانحدار يكون خطا افقيا، وبالتالي لا علاقة بين X و Y

بالنسبة (a): نرفض H_0 إذا كانت $t_c < -t_{n-2,1-\alpha}$ حيث $t_{n-2,1-\alpha}$ هو الفاصلة التي تترك على يسارها مساحة $1-\alpha$ تحت منحنى توزيع ستودنت بدرجة حرية $(n-2)$.

ملاحظة: نستطيع أن نعبر عن شكل المنطقة H_1 في الحالات الثلاثة كما يلي:

بالنسبة (b):

$$\begin{cases} H_0: B = b \\ H_1: B \neq b \end{cases}$$

نرفض الفرضية العدمية إذا كان:

$$t_c = \frac{B - b}{S_B} \rightarrow t_{n-2}$$

- إذا كان t_c يقع في المجال $\left[t_{n-2,\frac{\alpha}{2}} ; t_{n-2,1-\frac{\alpha}{2}}\right]$ نقبل H_0 ونرفض H_1.

- وبالتالي B يختلف عن b اختلافاً غير معنوي.

إذا كان t_c يقع خارج المجال $\left[t_{n-2,\frac{\alpha}{2}} ; t_{n-2,1-\frac{\alpha}{2}}\right]$ نقبل H_1 ونرفض H_0.

- وبالتالي B يختلف اختلافاً معنوياً عن b.

بالنسبة (c):

$$\begin{cases} H_0: B = b \\ H_1: B > b \end{cases}$$

نرفض الفرضية العدمية إذا كانت $t_c = \frac{B - b}{S_B} > t_{n-2,1-\alpha}$ حيث $t_{n-2,1-\alpha}$ هو الفاصلة التي تترك على يسارها مساحة $1-\alpha$ تحت منحنى توزيع ستودنت بدرجة حرية $(n-2)$. أما إذا كان $t_c = \frac{B - b}{S_B} < t_{n-2,1-\alpha}$ نقبل H_0 ونرفض H_1.

بالنسبة (d):

$$\begin{cases} H_0: B = b \\ H_1: B < b \end{cases}$$

نرفض الفرضية العدمية إذا كانت $t_c = \frac{B - b}{S_B} < -t_{n-2,1-\alpha}$ حيث $-t_{n-2,1-\alpha}$ هو الفاصلة التي تترك على يسارها مساحة $1-\alpha$ تحت منحنى توزيع ستودنت بدرجة حرية $(n-2)$ ونقبل H_0 ونرفض H_1 إذا كانت $t_c = \frac{B - b}{S_B} > -t_{n-2,1-\alpha}$.

(13):

ار المعنوية الكلية للنموذج:

ن اختبار معنوية النموذج الكلية باستخدام توزيع فيشر
وذلك كما يلي: . النموذج غير معنوي (الفرضية الصفرية)
. النموذج معنوي (مثالي): الفرضية البديلة

ولمعرفة التعبير عن فاحصة الفرضيتين، باختبار كما يلي:

$$\begin{cases} H_0 : B = \alpha = 0 \\ H_1 : B \neq 0 \end{cases}$$

ومن أجل اختبار فاحصة الفرضيتين نقوم بحساب القيمة الحصابية لفيشر، مقارنتها بقيمة الحصابية الجدولية المستخرجة من جدول فيشر، وذلك عند مستوى معنوية α ودرجة حرية $(1, n-2)$ (بالنسبة للنموذج الخطي البسيط) حيث

$$F_c = \frac{RSS/1}{ESS/n-2} = \frac{R^2/1}{1-R^2/n-2} = \frac{\hat{B}^2 \sum (x_i - \bar{x})^2 / 1}{\sum \hat{\varepsilon}_i^2 / n-2}$$

حيث إذا كان:

* $F_c > F(1, n-2, \alpha)$ ⟹ الفرضية H_0 مرفوضة، H_1 مقبولة
أي: B له معنوية حصابية، ومنه النموذج مثالي
أو العلاقة لها تمثيل.

* $F_c < F(1, n-2, \alpha)$ ⟹ H_0 مقبولة، H_1 مرفوضة، وبالتالي
B ليس معنويا حصابيا، ومنه النموذج غير مثالي
أو العلاقة ليست تمثيلية.

Scanned with Fast Scan

مثال: بالرجوع السلسلة السابقة (1)

لدينا: معادلة خط الانحدار المقدرة $\hat{C}_i = 2,671 + 0,687 R_i$ ، $n = 12$

R_i	C_i	$\hat{C}_i = 2,671 + 0,687 R_i$	$\hat{\varepsilon}_i = C_i - \hat{C}_i$	$\hat{\varepsilon}_i^2$
10	8	9,541	-1,541	$(-1,541)^2$
15	12	12,988	-0,988	$(-0,988)^2$
20	16	16,427	-0,427	$(-0,427)^2$
25	23	19,866	3,134	$(3,134)^2$
⋮	⋮	⋮	⋮	⋮
70	45	50,817	-5,817	$(-5,817)^2$
455	345		0	116,152

حساب التباين ومقدر الخطأ:

$$\hat{\sigma}^2_{\varepsilon} = \frac{\sum \hat{\varepsilon}_i^2}{n-2} = \frac{116,152}{10} = 11,6152$$

الخطأ المعياري للخطأ:

$$\hat{\sigma}_{\varepsilon} = \sqrt{\hat{\sigma}^2_{\varepsilon}} = \sqrt{11,615} = 3,408$$

مجال الثقة لتباين الخطأ: عند $\alpha = 5\%$

$$\sigma^2_{\varepsilon_i} \in \left[\frac{(n-2)\cdot \hat{\sigma}^2_{\varepsilon}}{\chi^2_{(n-2);\frac{\alpha}{2}}} ; \frac{(n-2)\cdot \hat{\sigma}^2_{\varepsilon}}{\chi^2_{(n-2),1-\frac{\alpha}{2}}} \right]$$

$\alpha = 0,05$ ، $(n-2) = (12-2) = 10$

$\chi^2_{(n-2),\frac{\alpha}{2}} = 20,48$ ($n-2 = 10$، $\frac{\alpha}{2} = 0,025$)

$\chi^2_{(n-2),1-\frac{\alpha}{2}} = 3,25$ ($n-2 = 10$، $1-\frac{\alpha}{2} = 0,975$)

$$\sigma^2_{\varepsilon_i} \in \left[\frac{10 \cdot 11,615}{20,48} ; \frac{10 \cdot 11,615}{3,25} \right]$$ ومنه،

$$\sigma^2_{\varepsilon_i} \in [5,671 ; 35,738]$$

Scanned with Fast Scan

(15) حال الثقة لكل من B و a عند مستوى $\alpha = 5\%$ ؛ تباين $\hat{a}$ و $\hat{B}$

$$\hat{\sigma}^2_{\hat{B}} = \frac{\hat{\sigma}^2_{\hat{\varepsilon}_i}}{\sum (x_i - \bar{x})^2} = \frac{11,611}{3872,916} = 0,0029$$

$$\hat{\sigma}^2_{\hat{a}} = \left[\frac{1}{n} \cdot \hat{\sigma}^2_{\hat{\varepsilon}_i} + \bar{x}^2 \cdot \hat{\sigma}^2_{\hat{B}}\right] = \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}\right] \hat{\sigma}^2_{\hat{\varepsilon}_i}$$

$$\hat{\sigma}^2_{\hat{a}} = \frac{1}{12} \cdot 11,611 + (37,917)^2 \cdot 0,0029 = 5,279$$

حساب مجال الثقة لكل من $\hat{a}$ و $\hat{B}$:

$$S_{\hat{B}} = \sqrt{\hat{\sigma}^2_{\hat{B}}} = \sqrt{0,0029} = 0,0546$$

$$S_{\hat{a}} = \sqrt{\hat{\sigma}^2_{\hat{a}}} = \sqrt{5,279} = 2,297$$

حساب مجال الثقة لـ B:

$$B \in \left[\hat{B} - S_{\hat{B}} \cdot t_{n-2,1-\frac{\alpha}{2}} ; \hat{B} + S_{\hat{B}} \cdot t_{n-2,1-\frac{\alpha}{2}}\right]$$

$$B \in [0,687 - 0,0546 \cdot 2,228 ; 0,687 + 0,0546 \cdot 2,228]$$ ومنه:

$$B \in [0,565 ; 0,808]$$

حساب مجال الثقة لـ a:

$$a \in \left[\hat{a} - S_{\hat{a}} \cdot t_{n-2,1-\frac{\alpha}{2}} ; \hat{a} + S_{\hat{a}} \cdot t_{n-2,1-\frac{\alpha}{2}}\right]$$

$$a \in [2,671 - 2,297 \cdot 2,228 ; 2,671 + 2,297 \cdot 2,228]$$

$$a \in [-2,446 ; 7,788]$$

Scanned with Fast Scan

اختبار معنوية المعالم، معنوية النموذج عند مستوى 5%

1- اختبار معنوية المعالم:

بالنسبة لـ β: نضع الفرضيات

$$\begin{cases} H_0: \beta = 0 \\ H_1: \beta \neq 0 \end{cases}$$

حساب الاحصائية المستعملة لستودنت t_c حيث

$$t_c = \frac{\hat{\beta} - \beta}{S_{\hat{\beta}}} \rightarrow t(n-2)$$

تحت الفرضية العدمية تكون:

$$t_c = \frac{\hat{\beta}}{S_{\hat{\beta}}} = \frac{0.6870}{0.0546} = 12,582$$

بما أن: $t_c > t_{tab} = t_{(n-2;1-\frac{\alpha}{2})} = 2.228$

نلاحظ أن $|t_c| > t_{tab}$ ← نرفض H_0 ونقبل H_1 أي أن β معلمة معنوية احصائيا (أي تختلف عن الصفر)

بالنسبة لـ $\hat{a}$: نضع الفرضيات

$$\begin{cases} H_0: a = 0 \\ H_1: a \neq 0 \end{cases}$$

حساب احصائية ستودنت t_c: حيث

$$t_c = \frac{\hat{a} - a}{S_{\hat{a}}} \rightarrow t_{n-2}$$

تحت الفرضية العدمية

$$t_c = \frac{\hat{a}}{S_{\hat{a}}} = \frac{2.671}{2,297} = 1,162$$

بما أن $t_{(n-2;1-\frac{\alpha}{2})} = 2.228$

نلاحظ أن $|t_c| < t_{tab}$ ← نقبل H_0 ونرفض H_1 أي: المعلمة a غير معنوية احصائيا [أي أن a لا تختلف عن الصفر]

2- اختبار المعنوية الكلية للنموذج:

نضع الفرضيات كما يلي: (النموذج غير معنوي)

$$\begin{cases} H_0: B = a = 0 \\ H_1: B \neq 0 \end{cases} \rightarrow$$ (النموذج معنوي)

من أجل القيام باختبار الفرضيات نقوم بحساب الفيشر وهذا ما يتطلب القيم الجدولية المستخرجة من جدول فيشر

(17)

$$F_c = \frac{R.SS/1}{ESS/n-2} = \frac{R^2/1}{\varepsilon \hat{\varepsilon}_i^2/n-2} = \frac{\hat{\beta}^2.(x_i-\bar{x})^2/1}{\varepsilon \hat{\varepsilon}_i^2/n-2}$$

$$= \frac{(0,687)^2.3872,916/1}{116,1^2/10} = \boxed{157,37 = F_c}$$

$$F_{(1,n-2),\alpha\%} = 4,96$$

F_{tab} $\{157,37 = F_c$ نقارن قيمة F_c المحسوبة بقيمة F_{tab} الجدولية، وعليه فالنموذج كلي معنوي أي المعادلة المقترحة لشرح العلاقة بين R و BG معادلة ذات مثيل جيد.

ملاحظة هامة: في حالة النموذج البسيط فقط: يكون

F_c (المحسوبة) $= (t_c)^2$ (B) الخاصة بالميل الحدي للادخار أي

$$F_{tab} = t_{tab}^2$$

$$158,2^1 = (12,582)^2 = (t_c)^2 \Rightarrow F_c = 157,37$$

$$4,963 = (2,228)^2 = t_{tab}^2 \Rightarrow F_{tab} = 4,96$$

صلاحية معامل التحديد: R^2

$$R^2 = \frac{RSS}{TSS} = \frac{\varepsilon(\hat{y}_i-\bar{y})^2}{\varepsilon(y_i-\bar{y})^2} = \frac{\hat{\beta}^2 \varepsilon(x_i-\bar{x})^2}{\varepsilon(y_i-\bar{y})^2} = \frac{\hat{\beta}^2.\varepsilon(x_i-\bar{x})^2}{\varepsilon y_i^2 - n\bar{y}^2}$$

$$R^2 = \frac{(0,687)^2.3872,916}{1948,2^1} = 0,9382 = 93,82\%$$

$(y_i-\bar{y})$	$(y_i-\bar{y})^2$
-29,75	$(-29,75)^2$
-16,75	$(-16,75)^2$
-12,75	$(12,75)^2$
⋮	⋮
16,25	$(16,25)^2$

$\varepsilon(y_i-\bar{y})^2 = 1948,2^1$

من هنا نقول أن R^2 تساوي 93,82% من التغيرات الحاصلة في BG تفسر من طرف التغيرات في R (المتغير المستقل) أما الباقي 6,18% فيرجع إلى عوامل أخرى لم يتضمنها النموذج (المتغير العشوائي).

Scanned with Fast Scan