مفهوم نموذج الانحدار الخطي البسيط المحاضرة الأولى

يعتبر الانحدار الخطي البسيط من الأساليب الإحصائية التي تستخدم في قياس العلاقة بين متغيرين على هيئة علاقة دالة، يسمى أحد المتغيرات (متغير تابع) والآخر (متغير مستقل أو مُفسِر) وهو المتسبب في تغير المتغير التابع، والانحدار الخطي كأداة للقياس لا تُحدد أي المتغيرات يكون تابع أو مستقل إنما يلجأ الباحث إلى النظرية الاقتصادية في تحديد المتغيرات، مثال ذلك: تفسير ظاهرة الاستهلاك بالدخل (مع ثبات العوامل الأخرى) فالنظرية الاقتصادية تقول إن استهلاك الفرد مرتبط بالدخل. وبالتالي فالباحث يسعى إلى إعطاء شكل للعلاقة بين المتغيرات الاقتصادية على شكل دالة:

$$Y=F(X)$$

حيث أن: -

Y : يمثل المتغير التابع (الاستهلاك)

X : يمثل المتغير المستقل (الدخل)

F : الدالة.

يمكن أن تأخذ الدالة أشكالا مختلفة قد تكون خطية، لوغاريتمية، أو أسية ... الخ، وغالبا ما يمكن تحويل أي نموذج إلى النموذج الخطي. ومن جانب اخر فان نماذج الانحدار يمكن تقسيمها الى عدة انواع، فهنالك الانحدار الخطي والانحدار الغير الخطي والانحدار البسيط والمتعدد. وتتحدد درجة الخطية على اساس درجة العلاقة المراد قياسها ففي حالة الانحدار الخطي تكون المعادلة الممثلة للعلاقة من الدرجة الاولى، وفي حالة الانحدار الغير الخطي تكون المعادلة الممثلة للعلاقة من الدرجة غير الاولى. اما عن صفتي بسيط ومتعدد فانهما يتحددان بعدد المتغيرات التفسيرية (المستقلة) التي تحتوي عليها معادلة الانحدار، فالانحدار البسيط يقيس العلاقة بين متغيرين أحدهما مستقل والاخر تابع. اما الانحدار المتعدد فهو يقيس العلاقة بين المتغير التابع وأكثر من متغير مستقل.

توصيف النموذج Model specification

تتمثل المرحلة الأولى من مراحل البحث في القياس الاقتصادي في توصيف النموذج محل الدراسة، وذلك من خلال التعبير عن العلاقات الاقتصادية المصاغة لفظيا في صورة رياضية حتى يمكن تقدير معالمها باستخدام طرق قياسية مختلفة.

ويعتمد الباحث في توصيفه للنموذج على ما يلي:

- مقترحات النظرية الاقتصادية
- اثباتات الدراسات التطبيقية التي تناولت المجال الذي يبحث فيه الباحث بوجه عام.
- المتوفر من بيانات حول الظاهرة محل الدراسة بوجه خاص

وتنطوي هذه المرحلة على ما يلي:

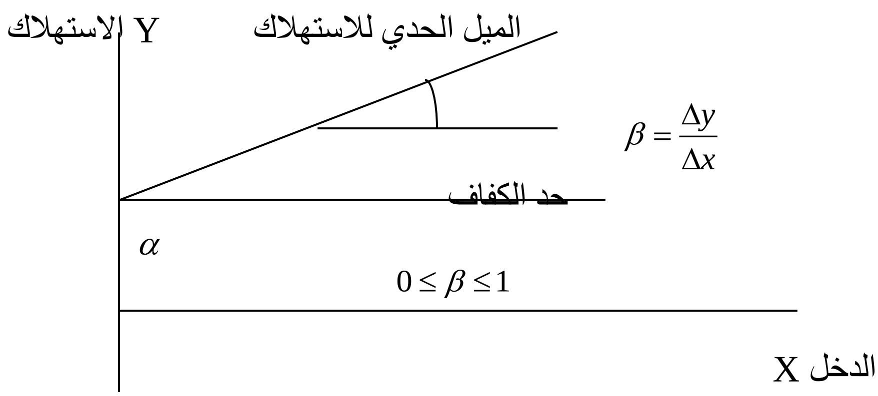
أولاً: اختيار متغيرات النموذج

سوف نفترض ان الهدف هو دراسة علاقة واحدة بين متغير تابع (Y) Dependent variable ومتغير مستقل (X) Independent variable. ويوجد هنالك العديد من الأمثلة التي يمكن تطبيق نموذج الانحدار الخطي البسيط عليها نذكر منها:

نموذج دالة الاستهلاك: -

يوضح الشكل (1) نموذج لدالة الاستهلاك، ويلاحظ بان α تمثل الحد الادنى للإنفاق الاستهلاكي الذي لابد ان يقوم به المجتمع في الفترة القصيرة حتى إذا انخفض الدخل المتاح الى الصفر. ويسمى بحد كفاف المجتمع. اما بالنسبة الى β فهي تمثل الميل الحدي للاستهلاك الذي يجب ان يكون بين الواحد والصفر. وهي تشير الى مقدار التغير في المتغير التابع (الاستهلاك) نتيجة لتغير المتغير المستقل (الدخل) بوحدة واحدة ويمكن التعبير عنها رياضيا وذلك من خلال الصيغة التالية: -

$$b = \frac{\partial y}{\partial x}$$



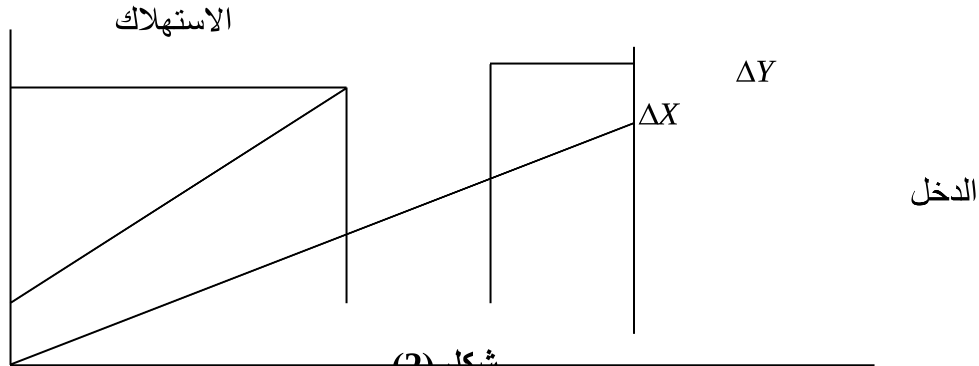
شكل (1)

نموذج دالة الاستهلاك

واما بالنسبة للميل المتوسط للاستهلاك، فانه يتم قياسه باستخدام المعامل المتوسط والذي يأخذ الصيغة التالية: -

$$\text{المعامل المتوسط} = \frac{\text{قيمةالمتغير التابع}}{\text{قيمةالمتغير المستقل}}$$

ويلاحظ ان المعامل المتوسط يمكن قياسه عند أي نقطة على خط الانحدار وذلك بميل الخط الواصل من هذه النقطة الى نقطة الاصل وكما موضح بالشكل التالي: -



شكل (2)

المعامل الحدي والمعامل المتوسط لدالة الاستهلاك

كما يلاحظ بان هنالك علاقة اقتصادية تربط بين الميل الحدي والميل المتوسط وتسمى بمرونة المتغير التابع للمتغير المستقل حيث تأخذ الصيغة التالية: -

$$\text{مرونة الاستهلاك للدخل} = \frac{\text{الميل الحدي للاستهلاك}}{\text{الميل المتوسط للاستهلاك}}$$

ثانياً: تحديد الصيغة الرياضية للعلاقة

تتعدد الصيغ الرياضية التي يمكن ان تعبر عن علاقة ما، ويمكن تحويل اغلب هذه الصيغ الى صيغة خطية، لان تلك الصيغة تعبر عن ابسط أنواع العلاقات. لذا فسوف نفترض ان العلاقة بين المتغير التابع Y والمتغير المستقل X هي علاقة خطية تأخذ الشكل التالي:

$$Y=B_0+B_1X \qquad \text{-------------- (1)}$$

حيث أن:

- y هو المتغير التابع (الذي يتأثر)
- x هو المتغير المستقل (الذي يؤثر)
- β_0 هو الجزء المقطوع من المحور الرأسي y ، وهو يعكس قيمة المتغير التابع في حالة انعدام قيمة المتغير المستقل x ، أي في حالة $x=0$
- β_1 ميل الخط المستقيم $(\beta_0+\beta_1 x)$ ، ويعكس مقدار التغير في y إذا تغيرت x بوحدة واحدة.

ثالثاً: تحديد شكل النموذج

سوف نفترض دراسة نموذج يحتوي على معادلة واحدة تربط بين متغيرين أحدهما تابع والأخر مستقل، وذلك بناء على ما افترضناه في الخطوتين السابقتين ويسمى عندئذ بالنموذج الخطي البسيط.

المحاضرة الثانية

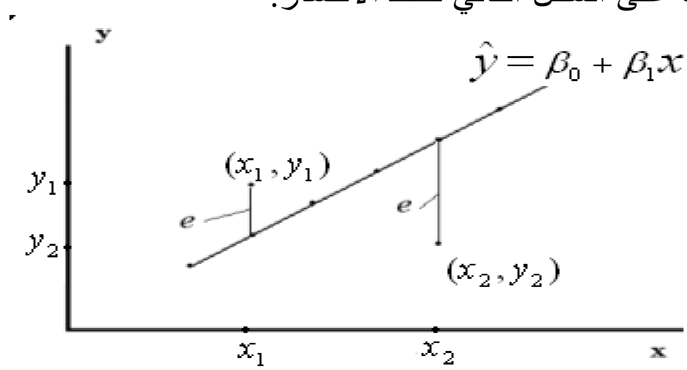
رابعاً: تحويل الصيغة الرياضية الى صيغة قياسية

نلاحظ ان الصيغة الرياضية للنموذج الخطي البسيط تحتوي على متغيرات منتظمة فقط، وهذا يعني اننا ننظر للعلاقة بين متغيرات النموذج باعتبارها علاقة تامة (غير احتمالية) حيث ان المتغير المستقل هو المتغير الوحيد المؤثر في المتغير التابع. ولكن في واقع الامر فان العلاقة ليست بهذه الصورة اذ يتعين ان يؤخذ في الحسبان أثر المتغيرات العشوائية بجانب المتغيرات المنتظمة. وذلك بإضافة حد عشوائي للصيغة الرياضية او ما يسمى بالخطأ العشوائي ويرمز له بالرمز e وبذا يتحول النموذج من نموذج رياضي يعكس علاقة تامة الى نموذج قياسي يعكس علاقة احتمالية. وبالتالي يمكن صياغة معادلة الانحدار الحقيقي للنموذج الخطي البسيط قياسيا على النحو التالي:

$$Y=B_0+B_1X+E \qquad \text{------------ (2)}$$

حيث أن:

E هو الخطأ العشوائي، والذي يعبر عن الفرق بين القيمة الفعلية y ، والقيمة المقدرة $\hat{y}=\beta_0+\beta_1 x$ ، أي أن : $e=y-(\beta_0+\beta_1 x)$ ، ويمكن توضيح هذا الخطأ على الشكل التالي لنقط الانتشار.



شكل (3)

الخطأ العشوائي في نموذج الانحدار البسيط

ويلاحظ من الشكل (3) اذا كان الخطأ العشوائي مساوي الى الصفر فان القيم المشاهدة سوف تنطبق على الخط المستقيم، بمعنى اننا نستطيع تفسير المتغير التابع Y تفسيرا كاملا بالأثر الخطي للمتغير المستقل X . وهذا الوضع يدفعنا للتساؤل حول الأسباب التي تؤدي الى انحراف القيم المشاهدة للمتغير التابع عن خط الانحدار والذي يؤدي لظهور الأخطاء العشوائية. والاجابة على هذا السؤال تكمن في الأسباب التالية:

1- **أخطاء توصيف الصيغة الرياضية للنموذج**

قد يقوم الباحث بتقدير صيغة للعلاقة غير الصيغة الحقيقية، كأن يقدر خطأ علاقة خطية وهي ليست خطية. ويترتب على ذلك اختلاف المعادلة المستخدمة عن المعادلة الحقيقية والتي تعبر عن الظاهرة محل الدراسة.

2- **اسقاط بعض المتغيرات الهامة في النموذج**

هناك بعض العوامل او المتغيرات التي يتعذر على الباحث قياسها كميا كالأذواق، او المتغيرات التي لا يمكن التنبؤ بحدوثها كالكوارث الطبيعية او المتغيرات المعروفة للباحث والقابلة للقياس الكمي، ولكن البيانات المتوفرة عنها غير كافية او غير دقيقة وبالتالي يقوم الباحث بإسقاطها ويترتب على ذلك اختلاف المعادلة المستخدمة عن المعادلة الحقيقية والتي تعبر عن الظاهرة محل الدراسة.

3- الأخطاء الراجعة للتجميع

نجد على سبيل المثال ان البيانات التجميعية للاستهلاك الكلي والدخل الكلي تعبر عن قيم متعلقة باستهلاك ودخول الافراد دون ان تعكس هذه النوعية من البيانات الخاصة بهيكل التوزيع للاستهلاك والدخل في المجتمع المدروس. فنجد ان تساوي مجموع الدخل القومي لبلدين لا يعني تساوي مستوى الرفاهية الاقتصادية فيهما، نظرا لإمكانية اختلاف توزيع الدخول بين الافراد في كلا البلدين. وبالتالي فان هذه الاختلافات تؤثر على الظاهرة محل البحث واسقاطها يترتب عليه خطا في المعادلة.

4- وجود خطأ في القياس

ويقصد بذلك حدوث أخطاء عند قياس المتغيرات المكونة للنموذج محل الدراسة عند اخذ المشاهدات. وتنتج هذه الأخطاء عند قيام الباحث بالمعالجة الإحصائية للبيانات، كما قد تنتج الأخطاء بسبب قصور في أساليب جمع البيانات. ومن ثم نجد ان الأخطاء السابقة تؤدي الى انحراف القيم المشاهدة عن الخط المستقيم والذي يمثله الخطأ العشوائي، مما يؤدي الى تحولها من علاقة تامة الى علاقة احتمالية بين متغيرات النموذج.

خامساً: التقدير Estimation

تتمثل المرحلة الثانية من مراحل البحث في القياس الاقتصادي في تقدير النموذج أي قياس القيم الرقمية لمعالم النموذج للوصول الى خط الانحدار الممثل للعلاقة بين متغيرات النموذج الذي تم توصيفه في المرحلة السابقة. ويعتمد الباحث أساساً في تقديره على ما يلي:

- بيانات واقعية تخص متغيرات النموذج
- طرق قياسية تستخدم في عملية التقدير

وتنطوي هذه المرحلة على ما يلي:

1. تحديد النموذج

سبق وان بينا ان نموذج الانحدار الخطي البسيط يمكن كتابته كما يلي:

$$Y=B_0+B_1X+E$$

ان هذا النموذج يعكس العلاقة الحقيقية بين متغيرات النموذج، وهذه المتغيرات يمكن الحصول عليها عند جمع بيانات واقعية عن كل القيم المشاهدة ويطلق على هذه المعادلة معادلة الانحدار

الحقيقي. الا ان خط الانحدار الحقيقي لا يكون معروفا على وجه التحديد وذلك لان النقاط التي نحصل عليها عادة ما تكون عينة صغيرة من مجتمع كبير أي انها لا تمثل سوى جزء صغير جدا من كل النقاط التي يمكن الحصول عليها، لذلك نقوم بتقدير خط الانحدار ونطلق عليه خط الانحدار المقدر ومعادلته:

$y=B_0+B_1x+e$

2. فروض النموذج Assumptions of the model

لكي يكون نموذج الانحدار الخطي البسيط قابلا للتقدير يتوجب توفر الفروض التالية:

- **(U_i)متغير عشوائي حقيقي.**

ان هذا الافتراض يعني ان (U_i) هو متغير حاله حال المتغيرات المستقلة الداخلة في الأنموذج ، الا انه استبعد من الانموذج المراد تقديره ، وهو بذلك يكون مع المتغيرات المستقلة تركيبة خطية تؤثر في المتغير التابع كما في الصيغة ( ) ، من جانب آخر ، فهو عشوائي وكل قيمة يأخذها في اية مدة زمنية ، تعتمد على المصادفة Depends on chance وذات تأثير متباين على (Y) او معدوم ولكل مشاهدة. وعندما نقول عشوائي يفترض انه استنزف من أي متغير احتواه ذي تأثير فعال على المتغير التابع (Y) ، وما يحتويه من متغيرات فإنها جاءت بطريقة غير عمدية ، اما اذا كان تحرك قيم احد او اغلب المتغيرات المحذوفة ، والمجتمعة بافتراض جدلي في (U_i) ، بنمط منتظم systematic pattern في الفترات المختلفة ، فان (U_i) لا يتمتع بافتراض العشوائية. ومع هذا نقول انه لا توجد صيغة اختبار عملية لذلك الافتراض لسبب بسيط هو ان (U_i) غير مشاهد، وحتى عندما يتم تقديره ويصبح رمزه (e_i) فان هذا التقدير اعتمد على احدى الطرائق المستخدمة في التقدير القائمة على افتراض العشوائية اصلاً، فكيف نستخدم مقدراً بني على افتراض العشوائية لاختبار ان كان هذا المتغير المقدر عشوائياً ام لا. من جانب آخر، فان (U_i) متغير حقيقي، لما يحتويه من متغيرات هي في الاصل حقيقية Real، وكان من المفترض ان تنظم للجزء التوضيحي في الانموذج، الا انه لأسباب كثيرة احتواها هذا المتغير تم استبعادها منه، ومن هنا نسلم بأهمية ان يكون (U) عشوائياً غير منتظم وخالٍ من المتغيرات ذات التأثير الفعال على (Y). وفي حالة كونه متغيراً مستقلاً غير عشوائي فان هذا يعني تحول العلاقة من علاقة قياسية الى علاقة رياضية.

- **متوسط القيمة المتوقعة للمتغير العشوائي يساوي صفراً.**

ان هذا الفرض يمكن صياغته رياضيا كما يلي: -

$E(U_i)=0$

والمقصود بهذه الفرضية ان قيم (U_i) المقابلة لقيم x_i والتي أخذت من المجتمع الإحصائي، احتوى هذا المجتمع على قيمها التي تكون أكبر، أصغر، مساوية للصفر، وان تلك القيم تتساوى في مجاميعها، فيكون متوسط (U_i) صفراً.

- **خاصية ثبات تجانس تباين الخطأ Homoscedsticity**

معنى هذه الفرضية ان تباين قيم (U_i) حول متوسط هذه القيم ثابتاً constant لجميع قيم (X) ، وهذا يعني ان قيم (U_i) سيكون لها التشتت Dispersion نفسه حول متوسطها لجميع قيم (X) . أي ان:

المحاضرة الثالثة

$Var.(U_i) = E(U_i)^2 = \sigma^2 U$

- **خاصية التوزيع الطبيعي للمتغير (U_i)**

المقصود بهذه الفرضية هو ان توزيع قيم (U_i) حول القيمة المتوقعة، والمفترض انها تساوي صفراً يكون على شكل جرس Bell-shaped symmetrical distribution لكل قيمة لـ (X).
ان هذا الافتراض يعني ضمنياً ان القيم الصغيرة لـ (U_I) تحدث باحتمالية أكبر من احتمالية حدوث القيم الكبيرة، ويعد هذا الافتراض من الافتراضات المهمة للغاية لبناء الاختبارات الاحصائية، وعمل قياسات لفترات الثقة. وعند عدم تحقق هذا الافتراض، وعلى الرغم من بقاء مقدرات المربعات الصغرى غير متحيزة ولها اصغر تبايناً ، الا ان تلك المقدرات لا تستوفي الصلاحيات الاحصائية لإجراء اختبارات مثل (χ^2 , F , T)) لكون هذه الاختبارات تبنى على التوزيع الطبيعي.

- **الاستقلالية المتسلسلة للأخطاء العشوائية serial Independence**

ان القيم المختلفة لـ (U_i) تكون مستقلة بعضها عن البعض الاخر، وان هذا يعني ان التباينات المشتركة لـ (U_i) مع (U_j) مساوية للصفر، ومن ثم فان قيمة المتغير العشوائي في مدة معينة، لا تعتمد على قيمته في مدة اخرى. ورياضياً فان هذه العلاقة تأخذ الصيغة الاتية:

$$Cov.(U_i U_j) = 0 \quad (i, j = 1,2.......n \quad , i \neq j)$$

ومن هذا يتبين ان هناك حالة انعدام للارتباط بين القيمة الجارية للمتغير المحذوف الواقع في (U) مع القيمة الماضية لنفس المتغير، وفي حالة وجود مخالفة لهذا الافتراض تظهر لنا مشكلة الارتباط الذاتي Auto-correlation.

- **قيم U_i غير مرتبطة باي من المتغيرات المستقلة**

المقصود بهذه الفرضية هو انعدام التباينات المشتركة بين (X's),(U's) وان توزيع قيم (X) مستقلاً عن توزيع قيم (U) أي ان :

$$Cov.(XU) = E\{[X_i - E(X_i)][U_i - E(U_i)]\} = 0$$

من هذا يتبين ان المصفوفة (X) ليست تصادفية، بمعنى انها تتوزع توزيعاً مستقلاً عن توزيع (U) ضمنياً هذا الافتراض يعني ان قيم (X) تبقى ثابتة Fixed وان الذي يتغير هو المتغير التابع (Y) نتيجة لاختلاف U ، بمعنى انه عندما نسحب عينات متتالية من المجتمع الاحصائي ، فان الذي يؤثر على الاختلاف في قيم (Y) هو المتغير العشوائي (U) فقط .

إضافة الى ذلك فان هناك افتراضات اخرى لا تقل اهمية عما سبقها، الا ان تلك الافتراضات مطلوب تحديدها قبل البدء بتقدير الانموذج، منها المتعلق بتشخيص Identification العلاقة المدروسة عندما يكون لدينا انموذج هيكلي يتكون من عدة معادلات، وآخر بعدم وجود اخطاء القياس Measurement Errors في المتغيرات المستقلة، وافتراض عدم وجود اخطاء التجميع Aggregation Errors التي قد يقع بها الباحث عند معالجة بياناته احصائياً. وافتراض صحة الصياغة Specification للأنموذج، والتي تمثل اولى واهم مراحل بناء الأنموذج والبحث في تحليل الانحدار.

3. **طريقة المربعات الصغرى الاعتيادية في التقدير** Ordinary least square

تعتمد طريقة المربعات الصغرى العادية على الحصول على مقدرات الانحدار بحيث يتم تصغير مجموع مربعات البواقي (Residuals) إلى أدنى قيمه لها.
ولتوضيح طرقة المربعات الصغرى الاعتيادية، نفرض لدينا نموذج الانحدار الخطي البسيط التالي: -

$$Y_i = \alpha + \beta X + u_i$$

ولغرض الحصول على مقدرات المربعات الصغرى العادية يجب أولا ايجاد مجموع مربعات البواقي $\sum u^2$ وكما يلي: -

$$\sum u_i^2 = \sum (Y_i - \hat{\alpha} - \hat{\beta} X)^2$$

وبعد ذلك يتم التوصل إلى الخط الذي تكون فيه مجموع مربعات البواقي أصغر ما يمكن [اختيار الخط الذي يدني مجموع مربعات البواقي إلى أصغر ما يمكن]. وذلك باستخدام التفاضل الجزئي بالنسبة للمجاهيل α, β حيث نحصل على :-

$$\frac{\partial\left(\sum u_i^2\right)}{\partial \hat{\alpha}} = (2)\sum (Y_i - \hat{\alpha} - \hat{\beta} X)(-1) \qquad \text{---------------} \ (3)$$

$$\frac{\partial\left(\sum u_i^2\right)}{\partial \hat{\beta}} = (2)(\sum (Y_i - \hat{\alpha} - \hat{\beta} X)(-X) \qquad \text{---------------} \ (4)$$

وبجعل المشتقة الاولى للمعادلة (3) مساوية الى الصفر: -

$$(-2)\sum (Y_i - \hat{\alpha} - \hat{\beta} X) = 0$$

$$\sum Y_i - \sum_i \hat{\alpha} - \hat{\beta} \sum_i X = 0$$

$$\sum_i \hat{\alpha} = \sum Y_i - \hat{\beta} \sum_i X$$

$$n\hat{\alpha} = \sum Y_i - \hat{\beta} \sum_i X$$

$$\hat{\alpha} = \frac{\sum Y_i}{n} - \hat{\beta} \frac{\sum_i X}{n}$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\,\overline{X} \qquad \text{------------------} \ (5)$$

وتسمى المعادلة (5) بالمعادلة الطبيعية الاولى.
وبمفاضلة المعادلة (4) بالنسبة للمعلمة β وجعل المشتقة مساوية الى الصفر، نحصل على: -

$$\frac{\partial\left(\sum u_i^2\right)}{\partial \hat{\beta}} = (2)(\sum (Y_i - \hat{\alpha} - \hat{\beta} X)(-X) = 0$$

$$\frac{\partial\left(\sum u_i^2\right)}{\partial \hat{\beta}} = -2\sum X (Y_i - \hat{\alpha} - \hat{\beta} X) = 0$$

$$(\sum X (Y_i - \hat{\alpha} - \hat{\beta} X) = 0$$

$$-\sum XY + \sum X \alpha + \sum \beta X^2 = 0$$

$$\sum XY = \alpha^{\wedge} \sum X + \beta^{\wedge} \sum X^2 \qquad \text{------------------} \ (6)$$

وتسمى المعادلة (6) بالمعادلة الطبيعية الثانية.
وبحل المعادلتين الطبيعيتين (5)، (6) نحصل على طرق قياسية لتقدير معالم النموذج. وكما يلي:

- ✓ نقوم بتعويض قيمة α " والتي تم الحصول عليها من المعادلة (5) " وذلك بالمعادلة (6) حيث نحصل على: -

$$\sum XY = \sum X\left(\frac{\sum Y}{n} - \beta\frac{\sum X}{n}\right) + \beta\sum X^2$$

- ✓ بالضرب في n ينتج ان :-

$$n\sum XY = \sum X\sum Y - \beta(\sum X)^2 + \beta\ n\sum X^2$$
$$n\sum XY - \sum X\sum Y = -\beta(\sum X)^2 + \beta\ n\sum X^2$$
$$= \beta\ n\sum X^2 - \beta(\sum X)^2$$
$$= \beta\left(n\sum X^2 - (\sum X)^2\right)$$

ولذلك يكون: -

$$\hat{\beta} = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum^{n} X_i \sum Y_t}{n\sum X_i^2 - \left(\sum X\right)^2}$$

وهنالك طريقة اخرى يمكن الحصول من خلالها على المقدرات وذلك باستخدام الانحرافات كما يلي:

$$\sum\ y_i = \sum(Y - \overline{Y})^2 \qquad = \sum Y_i^2 - n\overline{Y}^2$$
$$\sum xy = \sum(X - \overline{X})(Y - \overline{Y}) = \sum XY - n\overline{XY}$$
$$\sum x^2 = \sum(X - \overline{X})^2 \qquad = \sum X^2 - n\overline{X}$$

ولكن

$$\sum Y_i X_i = \sum X_i(\overline{Y}_i - \hat{\beta}\overline{X}) + \hat{\beta}\sum X_i^2$$
$$\sum Y_i X_i = n\overline{X}(\overline{Y} - \hat{\beta}\overline{X}) + \hat{\beta}\sum X_i^2$$
$$\sum Y_i X_i = n\overline{X}\,\overline{Y} - \hat{\beta}n\overline{X}\,\overline{X}) + \hat{\beta}\sum X_i^2$$
$$\sum Y_i X_i - n\overline{X}\,\overline{Y} = -\hat{\beta}n\overline{X}\,\overline{X} + \hat{\beta}\sum X_i^2$$
$$\sum XY - n\overline{XY} = -\hat{\beta}n\overline{X}^2 + \hat{\beta}\sum X_i^2$$
$$\sum xy = \beta\left[nX^2 - \sum X^2\right]$$
$$\sum xy = \beta\sum x^2$$
$$\beta = \frac{\sum xy}{\sum x^2}$$

المحاضرة الرابعة

4. خصائص المقدر الجيد

لقد تعرضنا في الموضوع السابق لطريقة المربعات الصغرى والتي تمكننا من قياس القيم المقدرة لمعالم النموذج الخطي البسيط من خلال بيانات العينة. وتهدف هذه الطريقة الى الحصول على المقدرات التي تجعل من مجموع مربعات البواقي أصغر ما يمكن، وللوصول الى هذا الهدف لا بد من توافر خصائص معينة لتلك المقدرات ويطلق عليها خصائص المقدر الجيد وهي:

- الخطية
- عدم التحيز
- الكفاءة
- الاتساق

وقبل ان نتعرض لخصائص المقدر الجيد سوف نقوم أولا بهذه التحويلات والتي سوف تساعدنا وتسهل علينا الكثير من الاثباتات القادمة.

يمكن كتابة $b=\frac{\sum x_i y_i}{\sum x_i^2}$ والتي تمثل تقدير معلمة الميل الحدي باستخدام طريقة المربعات الصغرى الاعتيادية بالصيغة التالية:

$$b=\sum \frac{x_i}{\sum x_i^2} y_i = \sum w_i y_i$$

بحيث ان:

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

وان قيم y،x مقاسة بالانحرافات حول المتوسط.

علما ان w_i تمتلك الخصائص التالية:

1. $\sum w_i = 0$
2. $\sum w_i^2 = \frac{1}{\sum x_i^2}$
3. $\sum w_i x_i = 1$

والان سوف ننتقل لاستعراض خصائص المقدر الجيد وهي:

1.4 الخطية Linearity

يعتبر المقدر خطي إذا كان يظهر كدالة خطية في القيم المشاهدة للمتغير التابع. ويمكن اثبات توافر هذه الخاصية في مقدرات المربعات الصغرى رياضيا كالتالي:

1.4.4 بالنسبة للمقدر $b^{\wedge}$

$$b^{\wedge}=\sum w_i y_i$$

وبما ان:

$$y = Y - \bar{Y}$$

لذلك فان:

$$\hat{b} = \sum w_i (Y - \bar{Y}) = \sum w_i * Y - \sum w_i * \bar{Y} = \sum w_i * Y$$

إذا فان b مقدر خطي في القيم المشاهدة للمتغير التابع Y باعتبارها دالة في المتغير التابع Y.

2.4.4 بالنسبة للمقدر $\hat{a}$

بما ان:

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

إذا يكون:

$$\hat{\alpha} = \frac{\sum Y_i}{n} - \bar{X}\sum w_i Y_i$$

وبأخذ Y_i عامل مشترك:

$$\hat{\alpha} = \sum (\frac{1}{n} - \bar{X} w_i) Y_i$$

اذا فان $\hat{\alpha}$ مقدر خطي في القيم المشاهدة للمتغير التابع Y

2.4 عدم التحيز Unbiasedness

ان مقدرات المربعات الصغرى الاعتيادية لكل من α، β غير متحيزة. وهذا يعني ان الفرق بين القيمة المتوقعة للمقدر والقيمة الحقيقية لمعلمة المجتمع الاصلية مساويا للصفر. أي بمعنى اخر:

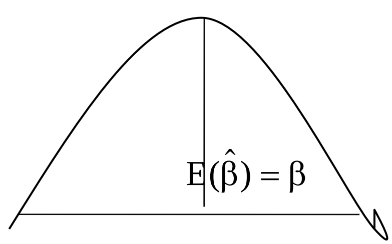
$$E(\hat{\alpha}) = \alpha$$

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

فاذا جمعت عينات كثيرة وفي كل عينه نحسب $\hat{\alpha}$ ومن ثم يتم أخذ المتوسط، ذلك المتوسط نظريا يجب أن يتساوى مع المعلمة الحقيقية.

ان هذه الأوضاع كلها نظريه بحتة، في الواقع لا يكون عندنا عدد من العينات اذ غالبا ما يتوفر عينه واحدة فقط وتعطينا قيمه واحدة $\hat{\alpha}$، وقيمه واحدة $\hat{\beta}$ يعتمد عليها في التحليل، من الناحية النظرية نقول أن هذه المقدرات يتوقع أنها تساوي القيمة الحقيقية ومن الناحية الأخرى ان القيمة الحقيقة لا نعرفها وبالتالي فان هذه الخصائص خصائص نظريه بحتة.

ولتوضيح خاصية عدم التحيز بيانيا يتم رسم دالة احتمال $\hat{\beta}$ ، ان خاصية عدم التحيز تقول أن توزيع احتمال $\hat{\beta}$ يأخذ الشكل ( )



يتمركز حول القيمة الحقيقية لـ β وهذا يعني أن القيمة المتوقعة لـ $\hat{\beta}$ تســاوي β $E(\hat{\beta}) = \beta$ وأن قيمة β تساوي المعلمة الحقيقية ونفس التحليل ينطبق على α

ولإثبات ان مقدرات المربعات الصغرى الاعتيادية تكون غير متحيزة نظريا يكون ذلك كما يلي:

- **بالنسبة للمقدرة b^**

بما ان:

$$b^{\wedge} = \sum w_i y_i$$

لذلك فان:

$$b^{\wedge} = \sum w_i(\alpha + \beta X + U) = b + \sum w_i U$$

بحيث ان:

$$\sum w_i = 0$$
$$\sum w_i x_i = 1$$

وبأخذ القيمة المتوقعة للطرفين:

$$E\ b^{\wedge} = b + \sum w_i EU$$

حيث ان:

$$EU=0$$

لذلك فان:

$$E\ b^{\wedge} = b$$

ومعنى ذلك ان تقديرات المربعات الصغرى الاعتيادية لمعلمة الميل الحدي هي تقديرات غير متحيزة.

- **بالنسبة للمقدرة α^**

بما ان:

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

لذلك فان:

$$\hat{\alpha} = \frac{\sum y}{n} - \bar{X}\sum wY = \sum(\frac{1}{n} - \bar{X}w)Y = \sum(\frac{1}{n} - \bar{X}w)(\alpha + \beta X + U)$$

وبضرب الاقواس معا واخذ الخطأ العشوائي عامل مشترك ينتج ان:

$$\hat{\alpha} = \alpha + \sum(\frac{1}{n} - \bar{X}w) * U$$

وبأخذ القيمة المتوقعة للطرفين نحصل على:

$$E\hat{\alpha} = \alpha$$

لذلك فان تقديرات المربعات الصغرى الاعتيادية بالنسبة لمعلمة القاطع هي تقديرات غير متحيزة.

ويتعين إيجاد صيغ تباينات وتغايرات المعالم المقدرة قبل الدخول في باقي خصائص مقدرات المربعات الصغرى، نظرا لأهمية تلك المؤشرات الإحصائية في الاثباتات الخاصة بخاصيتي الاتساق والكفاءة وهي كما يلي:

المحاضرة الخامسة

- **تباين مقدرة القاطع:**

ان تبـاين اي قيمـة تتـوزع حـول وسط معين هو معدل تشتت هذه القيم عن الوسط الحسـابي لهـا ويكون القانون الخاص بتباين مقدرة القاطع:

$$V(\hat{\alpha}) = E\{\hat{\alpha} - E(\hat{\alpha})\}^2$$

وبإجراء بعض الخطوات يمكن برهنة تباين $\hat{\alpha}$ وكما يلي:

1- بما اننا اوجدنا في الخاصية الثانية ان:

$$\hat{\alpha} = \alpha + \sum(\frac{1}{n} - \bar{X}w) * U$$

2- بالتعويض عن $\hat{\alpha}$في معادلة $V(\hat{\alpha})$ نحصل على:

$$V(\hat{\alpha}) = E\left[\alpha + \sum\left(\frac{1}{n} - \bar{X}w_i\right)U_i - \alpha\right]^2 = E\left[\sum\left(\frac{1}{n} - \bar{X}w_i\right) * U_i\right]^2 = \sum E\left(\frac{1}{n} - \bar{X}w_i\right)^2 * U_i{}^2$$

3- بضرب الاقواس معا واجراء العمليات الحسابية نحصل على:

$$V(\hat{\alpha}) = V(U)\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i{}^2}\right]$$

4- بتوحيد المقام:

$$V(\hat{\alpha}) = V(U)\left[\frac{\sum x_i{}^2}{n\sum x_i{}^2} + \frac{n\bar{X}^2}{n\sum x_i{}^2}\right]$$

5- بما ان:

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

6- وبعد فتح الاقواس واجراء عملية التبسيط نحصل على:

$$\sum x_i{}^2 = \sum X_i{}^2 - n\bar{X}^2$$

7- وبتعويض النتيجة في الخطوة 6 في معادلة التباين بالخطوة 4 نحصل على:

$$V(\hat{\alpha}) = V(U)\left[\frac{\sum x_i{}^2}{n\sum x_i{}^2}\right]$$

والذي يمكن كتابته كما يلي:

$$V(\hat{\alpha}) = \frac{\sigma^2 \sum X^2}{n\sum x^2}$$

ان الصيغة (   ) تمثل تباين تقدير معلمة القاطع، كما توجد صيغة أخرى للتباين وهي:

$$V(\hat{\alpha}) = \sigma^2\{\frac{1}{n} + \frac{\bar{X}^2}{\sum x^2}\}$$

- **تباين المعلمة المقدرة للميل الحدي:**

ان المقصود بتباين $b^{\wedge}$ ذلك التباين ما بين القيمة المقدرة للمعلمة $b^{\wedge}$ لكل عينة من العينات باستخدام المقدر $b^{\wedge}$ والقيمة الحقيقية لمعلمة المجتمع b الاصلية ومن ثم يمكن إيجاد الصيغة الرياضية لتباين $b^{\wedge}$ كالتالي:

$$V(\hat{\beta}) = \{\hat{\beta} - E(\hat{\beta})\}^2$$

بما اننا اوجدنا في الخاصية الثانية ان:

$$b^{\wedge} = b + \sum w_i U$$

وبالتعويض عن $b^{\wedge}$ في معادلة $V(\hat{\beta})$ نحصل على:

$$v(b^{\wedge}) = E(b + \sum w_i U - b)^2 = E(\sum w_i U)^2 = E(w_1 U_1 + \ldots\ldots\ldots\ldots + w_n U_n)^2$$

وبعد عملية فتح الاقواس وإدخال التربيع نحصل على:

$$v(b^{\wedge}) = \sum w_i^2 EU^2 + \sum\sum w_i w_j EU_i U_j = V(U)\sum w_i^2$$

لذلك فان التباين الخاص بـ $\hat{\beta}$ هو:

$$V(\hat{\beta}) = \frac{\sigma^2}{\sum x^2}$$

- **التباين المشترك بين المعالم المقدرة:**

يمكن إيجاد التباين المشترك لمعالم نموذج الانحدار المقدر وذلك كما يلي:

$$\mathrm{Cov}(\hat{\alpha}, \hat{\beta}) = E\left[\{\hat{\alpha} - E\hat{\alpha}\}\{\hat{b} - E\hat{b}\}\right] = E\left[\{\hat{\alpha} - \alpha\}\{\hat{b} - \beta\}\right]$$

وبما ان:

$$b^{\wedge} = b + \sum w_i U$$

$$\hat{\alpha} = \alpha + \sum(\frac{1}{n} - \bar{X}w) * U$$

وبالتعويض عن $\hat{\alpha}, \hat{b}$ في $\mathrm{Cov}(\hat{\alpha}, \hat{b})$:

$$\mathrm{Cov}(\hat{\alpha}, \hat{\beta}) = E\left[\left\{\alpha + \sum(\frac{1}{n} - \bar{X}w) * U - \alpha\right\}\left\{\beta + \sum w_i U - \beta\right\}\right]$$

$$\mathrm{Cov}(\hat{\alpha}, \hat{\beta}) = E\left[\left\{\sum(\frac{1}{n} - \bar{X}w) * U\right\}\left\{\sum w_i U\right\}\right]$$

$$\mathrm{Cov}(\hat{\alpha}, \hat{\beta}) = \sum E(\frac{1}{n} - \bar{X}w) * U * w_i * U = \sum(\frac{1}{n} - \bar{X}w) * w_i * EU^2$$

وبعد عملية فتح الاقواس والتبسيط نحصل على:

$$\mathrm{Cov}(\hat{\alpha}, \hat{\beta}) = -V(U) * \frac{\bar{X}}{\sum x_i^2} = -\sigma^2 * \frac{\bar{X}}{\sum x_i^2}$$

3.4 الاتساق Consistency

يعتبر المقدر متسقا لمعلمة المجتمع الاصلية إذا توفرت فيه ما يلي:

- خاصية عدم التحيز

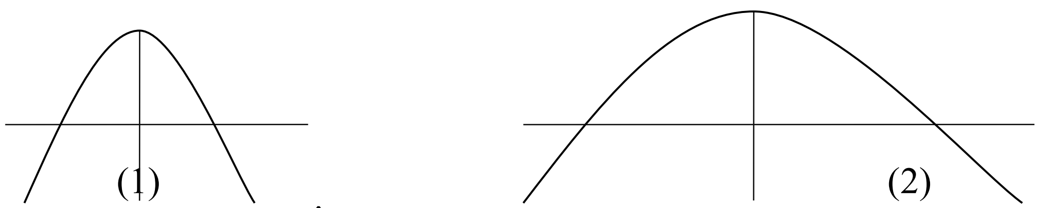
- اقتراب تباين المعلمة المقدرة من الصفر عندما يؤول حجم العينة n الى ما لا نهاية.
ويمكن اثبات توافر خاصية الاتساق في مقدرات المربعات الصغرى رياضيا كالتالي:
- بالنسبة لمقدر معلمة الميل الحدي
بما ان تباين المقدر $b^{\wedge}$ يأخذ الصيغة الرياضية التالية:

$$V(\hat{\beta}) = \frac{\sigma^2}{\sum x^2}$$

4- الكفاءة (اقل تباين) Efficiency

الخاصية الثالثة لمقدرات م ص ع تمتلك آدني تباين هذه الخاصية لها أهمية بالغة في الاقتصاد القياسي لان آدني تباين يعتبر مؤشر إلى دقة القياسات، آدني تباين يعني أعلى دقة من ناحية القياسات.

هناك علاقة عكسية بين التباين ودقة القياسات كلما زاد التباين كلما انخفضت دقة القياسات وكلما قل ارتفعت دقة القياسات. لأن مقدرات م ص ع $\hat{\alpha}$ $\hat{\beta}$ تلك المقدرات تمتلك آدني تباين نعني مقارنه بمقدرات أخرى تقاس بطريقه مختلفة عن م ص ع فان مقدرات م ص ع تمتلك آدني تباين إي تتحلى بأعلى دقه. نفترض إن هناك مقدرات لـ α β تحصل عليها بطريقه مختلفة ونفترض إن المقدرات الأخرى α'', β'' ا ذا افترضنا أن تلك المقدرتين خطيه وغير متحيزة سيكون الاختلاف في خاصية أن مقدرات م ص ع $\hat{\alpha}$ $\hat{\beta}$ تمتلك أعلى دقة.



$$E(\hat{\beta}) = \beta \qquad\qquad E(\beta'') = \beta$$

$$V(\hat{\beta}) < V(\beta'')$$

في الحالة (1) استخدمت مقدرات م ص ع . في الحالة الثانية (2) مقدرات أخرى غير م ص ع ، في الشكل التوزيع الاحتمالي لقيمة المقدرات β'', $\hat{\beta}$. في (1) يتبين ان التباين قليل، درجة الانتشار لـ $\hat{\beta}$ اقل وبالتالي تتمركز قيم $\hat{\beta}$ حول القيمة الحقيقية وفي الشكل (2) قد نحصل على قيم حول β لكنها بعيده عن المعلمة الحقيقية.

من الشكل إن احتمال الحصول على $\hat{\beta}$ أقرب للمعلمة الحقيقية من β'', وبالتالي درجة احتمال العثور على $\hat{\beta}$ أقرب مما سواها، هذا ما يقصد بخاصية أدنى تباين.

من النتائج التي توصلنا أليها عن مقدرات م ص ع يمكن أن نقول أن شـــــكل التوزيع الاحتمالي الخاص بالمقدرات $\hat{\alpha}$ $\hat{\beta}$

$$\hat{\alpha} \sim N\left[\alpha, \sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x^2}\right)\right], \quad \hat{\beta} \sim N\left[\beta, \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}\right]$$

من المعادلتين يتبين انه :

1- كلما زاد التباين σ^2 كلما زاد تباين المقدرات $\hat{\alpha}$ $\hat{\beta}$.

2-كلما كان انتشار قيم X اكبر كلما قل تباين $\hat{\alpha}$ $\hat{\beta}$