2017/4/16 الاسبوع الرابع (3) ٣

القياس

الاختبارات الاحصائية لمعنوية المعلمات المقدرة بطريقة (OLS) المربعات الصغرى

Statistical Tests of Significance of the (OLS) Estimats

ان اهم الاختبارات الاحصائية والاكثر استعمالاً من الاقتصاد القياسي هما اختبارين اساسيين Correlation Coefficient Squares

الاختبار الاول: وهو مربع معامل الارتباط (r^2) والذي يستخدم للحكم على القابلية التوضيحية لخط الانحدار بين المتغير المعتمد (Y) والمتغير التوضيحي (X)

الاختبار الثاني: يعتمد على الاخطاء المعيارية (Standard Errors) للمعلمات المقدرة، وتطبيق معادلة درجة الحكم على الدلالة الاحصائية من المعلمات المقدرة ... ان هذا الاختبار يعطينا مقياس لدرجة الثقة التي يمكن ان نعطيها للتقديرات كما ان هذا الاختبار يمكن الباحثين بقرار جودة التقديرات للتعبير عن المعلمات الحقيقية للمجتمع الاحصائي.

ومن هذه الاختبارات

① - معامل التحديد R^2 (مربع معامل الارتباط r^2) Coefficient of Determination

بعد تقدير النموذج القياسي فان الباحث يرغب في معرفة كم تقترب المشاهدات الحقيقية من خط الانحدار المقدر وتسمى هذا الاقتراب بجودة التوفيق (Goodness of fit)

ان مقياس هذه الجودة هو معامل التحديد R^2 ويسمى احياناً معامل التفسير اي انه يفسر كم من التغيرات في المتغير التابع (Y) يتم توضيحها عن طريق المتغيرات المستقلة التوضيحية في النموذج القياسي. ويطلق عليه ايضاً تسمية معامل التوضيح.

$$R^2 = \frac{\Sigma \hat{y}^2}{\Sigma y^2} = \frac{\text{التباين الموضح}}{\text{التباين الكلي}} = \frac{\text{Variation Explained}}{\text{Total Variation}}$$

وهذا يعني اساساً ان تقيس انتشار او تشتت المشاهدات (dispersion of Observations) حول خط الانحدار

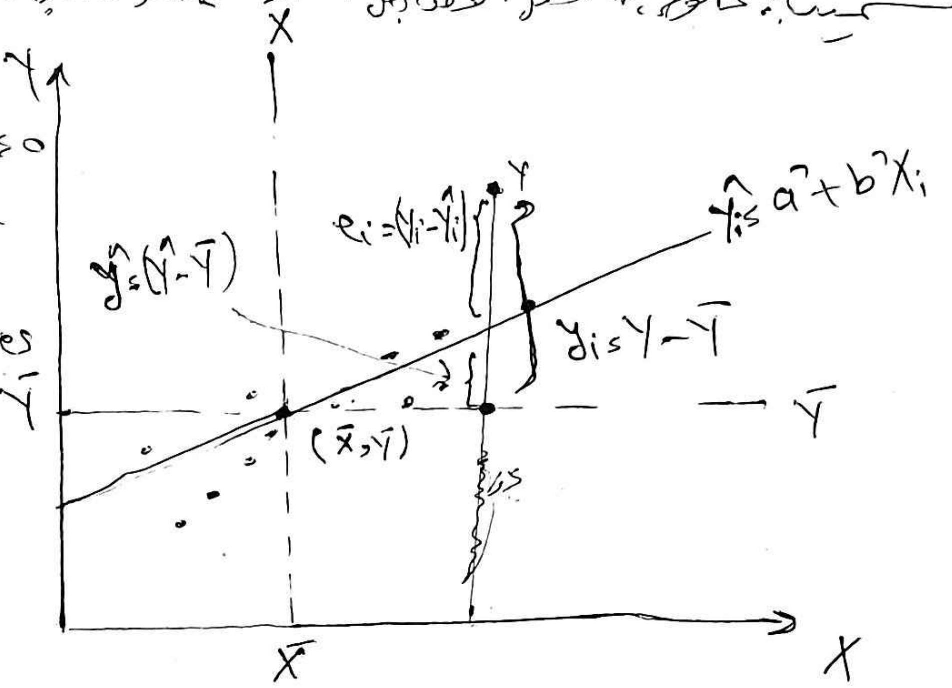
نلاحظ من الشكل ان هناك نقاط انتشار مبعثرة ولكي نعرف جودة التوفيق ونوضح ماذا تعني نقطة معينة من هذه النقاط ولتكن (Y)، نلاحظ ان التباين الكلي اي ابتعاد النقطة (Y) عن وسطها الحسابي ($\bar{Y}$) هو عبارة عن (y) وهذا التباين الكلي يقسم الى قسمين كما هو في الشكل، ومن اجل الحصول على القيمة الكلية (Y) نحتاج الى تربيع الانحرافات

$$\sum (Y-\bar{Y}) = \sum y_i = 0$$

$$TSS = \sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} (Y-\bar{Y})^2$$

Total Sum squares

TSS



القسم الاول هو ما تم توضيحه عن طريق خط الانحدار ($\hat{Y}_i = \hat{a} + \hat{b} X_i$) وهو عبارة عن ($\hat{Y}$) ويسمى بالتباين الموضح (RSS) وهذه الانحرافات تمثل مجموع الانحرافات في المتغير التابع الذي تم توضيحه بخط الانحدار

$$\sum_{i=1}^{n} \hat{y}^2 = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 = RSS$$ Regression Sum Squares

مجموع مربعات الانحدار

Explained Variation

اما القسم الثاني فهو التباين غير الموضح والذي لا يوضحه خط الانحدار والذي يعود الى الاضطراب الذي يسببه المتغير العشوائي (U_i) وعليه فان مجموع مربعات الانحرافات تعطينا مجموع الانحرافات في المتغير التابع (Y) الذي لم يوضحها خط الانحدار ESS

Error Sum squares

Unexplained variation

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y})^2$$

اي ان

$$TSS = RSS + ESS$$

$$\sum y^2 = \sum \hat{y}^2 + \sum e_i^2$$

(٤)

- ويتبين لنا من خلال (r^2) أكبر نسبة التقيدات في Y التي هي خرجي X، لذلك يطلق عليه معامل التحديد أو التقرير، فإذا وجدنا على سبيل المثال ($R^2 = 0.90$) فإن ذلك يعني أن خط الانحدار يعطينا توفيقاً جيداً للبيانات المشاهدة لأنه يوضح 90% من التقيدات في قيم (Y) حول متوسطها، أما النسبة 10% الباقية فترجع إلى المتغير العشوائي وما يحويه من عوامل.

- إن قيمة معامل التحديد r^2 تتراوح بين الصفر والواحد الصحيح، أي أن

$$0 \leq r^2 \leq 1$$

وهناك صيغ رياضية مختصرة لاحتساب r^2 هي

$$R^2 = \frac{(\sum y x_i)^2}{\sum x_i^2 \sum y^2} = \frac{\hat{\beta} \sum y x}{\sum y_i^2} = \frac{\hat{\beta}^2 \sum x_i^2}{\sum y^2} = 1 - \frac{\sum e_i^2}{\sum y^2}$$

② اختبار الأخطاء المعيارية للمقدرات Standard Errors

- إن اختبار الأخطاء المعيارية للتقدير شائع الاعتماد القياسي التطبيقي، وهذا الاختبار يساعد في تحديد ما إذا كانت التقديرات $\hat{a}, \hat{b}$ تختلف عن الصفر اختلافاً معنوياً، أي ما إذا كانت العينة التي استخدمت للتوصل إلى هذه التقديرات مأخوذة من مجتمع إحصائي معاملاته الحقيقية مساوية للصفر ($a = b = 0$) وبعبارة أخرى عادل اختبار

Null hypothesis $H_0: \hat{a} = \hat{b} = 0$ فرضية العدم

Alternative hypothesis $H_1: \hat{a} \neq 0$ مقابل الفرض البديل

$\hat{b} \neq 0$

والاختبار يتم وفق الخطوات التالية: ① حساب الخطأ المعياري $\hat{a}, \hat{b}$

- استخدام صيغ التباين الخاصة بالتقديرات للمقدرات

$$S(\hat{a}) = \sqrt{var(\hat{a})}$$

$$S(\hat{b}) = \sqrt{var(\hat{b})}$$

(3) $Variance(\hat{a}) = \frac{\sigma_u^2 \sum X_i^2}{n \sum x_i^2}$

$var(\hat{b}) = \frac{\sigma_u^2}{\sum x_i^2}$

(4) نعرف ان قيمة الاخطاء المعيارية للمقدرات هذه مع نصف القيمة العددية للمقدرات $\hat{a}$ و $\hat{b}$ فاذا كانت الاخطاء المعيارية اقل من نصف القيمة العددية للمقدر اي ان

$$S(\hat{a}) < \frac{\hat{a}}{2} \quad , \quad S(\hat{b}) < \frac{\hat{b}}{2}$$

نستخلص من هذا ان التقدير ذو معنوية احصائية اي نرفض فرضية العدم ($\hat{a}=\hat{b}=0$) ونقبل بالفرض البديل ($\hat{a}\neq 0 , \hat{b}\neq 0$) وبالعكس اذا كانت الاخطاء المعيارية للتقدير اكبر من نصف القيمة العددية لهذا المقدر

$$S(\hat{a}) > \frac{\hat{a}}{2} \quad , \quad S(\hat{b}) > \frac{\hat{b}}{2}$$

فان التقديرات غير معنوية احصائيا اي نقبل بفرضية العدم ($\hat{a}=\hat{b}=0$)

المعنى الاقتصادي للاختبار

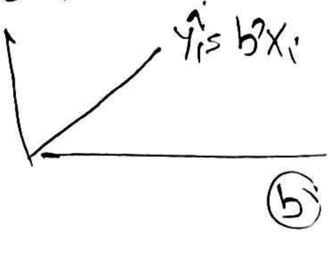
ان قبولنا او رفضنا لفرضية العدم له مدلولات اقتصادية واضحة، فقبولنا فرضية العدم القائلة بان معامل الانحدار الحقيقي B مساوي للصفر يعني ان المتغير التوضيحي (X) الذي يرتبط به هذا المعامل لا يؤثر في المتغير التابع Y وبالتالي فان المعروضات لا تتحول مع الدراسة. اي ان العلاقة بين Y و X [illegible] بالشكل الاتي

$$Y_i = a + (0)X_i$$

المعنى الهندسي: سبق ان ذكرنا ان ($\hat{a}$) هو الجزء المحصور بين نقطة تقاطع خط الانحدار مع المحور Y ونقطة الاصل، اما ($\hat{b}$) فيشير الى ميل الخط الانحدار وهنا:

(a) اذا كانت $S(\hat{a}) > \frac{\hat{a}}{2}$ فاننا نقبل بفرضية العدم ($\hat{a}=0$) هذا يعني ان خط الانحدار يمر بنقطة الاصل والعلاقة بين Y و X تأخذ الشكل الاتي

$$Y_i = bX_i + U_i$$



(b)

(35)

ⓑ - اما اذا وجدنا ان $\frac{\hat{b}}{2} < S(\hat{b})$ فانه نقبل بفرض العدم $\hat{b} = 0$ مما يشير الى ان العلاقة بين (X و Y) تتخذ الشكل الآتي $\hat{Y} = \hat{a} + u_i$ اي ان الخط الافقي مساويا صفرا بمعنى ان هذا الخط سوف يكون موازيا لمحور X من نقطة مساوية ($\hat{a}$)

② اختبار Z   Z test

يستند هذا الاختبار على التوزيع المعتدل المعياري ويتطلب استخدامه استيفاء احد الشروط التالية

① ان تكون القيمة الحقيقية لتباين المتغير العشوائي (σ_u^2) معروفة

② اذا لم تكن هذه القيمة معروفة فيكتفي بان يكون حجم العينة المستخدمة (n) كبيرا الى درجة كافية ($n > 30$)

فاذا لم يستوفى احد هذين الشرطين فان علينا استخدام الاختبار التالي وهو اختبار (t)

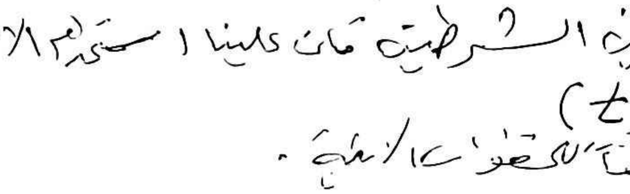
ويتم هذا الاختبار وفقا للخطوات الآتية.

① - تحديد كل من فرض العدم والفرض البديل

$H_0: \hat{a} = 0 , \hat{b} = 0$

$H_1: \hat{a} \neq 0 , \hat{b} \neq 0$

② - اختيار مستوى المعنوية الذي نقرر على اساسه قبولنا او رفضنا لفرضية العدم، ومن المتبع في الاقتصاد القياسي التطبيقي اختيار مستويات المعنوية (5% و 1%) ونظرا لعدم معرفتنا بالقيمة الفعلية لمعامل المتغير العشوائي فاننا نجري اختبار ذو جانبين بمعنى اننا نختار جانبي التوزيع المعتدل كمناطق رفض، حيث يشمل كل جانب على نصف احتمال مستوى المعنوية الذي وقع عليه الاختيار



وبالرجوع الى جدول التوزيع المعتدل المعياري نجد ان القيمة الحرجة للمتغير Z المقابلة لاحتمال (0.025) في كل طرف من المنحنى هي $Z = \mp 1.96$

③ - نستخرج قيمة Z^* المحسوبة

$$Z^*(\hat{a}) = \frac{\hat{a}}{S(\hat{a})} , \quad Z^*(\hat{b}) = \frac{\hat{b}}{S(\hat{b})}$$

36

5. المقارنة بين Z^* و Z الجدولية فاذا كانت

$$-Z < Z^* < +Z$$

نقبل بفرض العدم ونرفض الفرض البديل

فاذا كانت

$$-Z > Z^* > +Z$$

نرفض فرض العدم ونقبل الفرض البديل.

4- اختبار t — t test

كما اسلفنا ان اختبار Z يستخدم اذا توفر احد شرطين هما ان تتوفر لدينا معرفة بالقيمة الحقيقية لتباين المجتمع الاحصائي لفترة التقدير عند حجم العينة المستخدمة، اما اذا لم تتوفر لدينا القيمة الحقيقية للتباين وجدنا ان حجم العينة غير كبير بدرجة كافية ($n \geq 30$)، فانه نستخدم اختبار (t) لانه بالامكان في هذه الحالة استخدام تقديرات العينة للتباين كتقريب كافي لقيمة تباين المجتمع الاحصائي، ولغرض تطبيق اختبار t تتبع الخطوات التالية

1- تحديد فرض العدم والفرض البديل

$$H_0: \hat{a} = \hat{b} = 0$$

$$H_1: \hat{a} \neq 0, \hat{b} \neq 0$$

2- نستخرج قيمة t الجدولية بدرجة حرية $(n-k)$ ومستوى معنوية 5% و 1% من الجدول المسماه بتوزيع ستيودنت

$$t\left(\frac{\alpha}{2}, n-k\right), \quad t(\alpha, n-k)$$

3- نستخرج قيمة t^* المحتسبة باستخدام الصيغة

$$t^*(\hat{a}) = \frac{\hat{a}}{S(\hat{a})}, \quad t^*(\hat{b}) = \frac{\hat{b}}{S(\hat{b})}$$

4- نقارن قيمة t^* المحتسبة مع t الجدولية

$$t^*_{\text{المحتسبة}} > t_{\text{الجدولية}}$$

رفضنا فرض العدم وقبلنا الفرض البديل، اي ان التقديرات ذات معنوية احصائية

اما اذا كانت

$$t^*_{\text{المحتسبة}} < t$$

قبلنا بفرض العدم وبالتالي تكون التقديرات غير معنوية احصائياً

(37)

5 تقدير فترة المقدر حدي الثقة Confidence Intervals for a, b

قد يرغب الباحث عوضاً عن تقدير قيمة محددة (تقدير نقطة) للمعلمة a و b انت نقوم بتقدير حدي اصغر واكبر وحساب احتمال وقوع القيمة المجهولة بينهما، ويتم ذلك وفق الصيغ الرياضية التالية.

اي بتعبير اخر ان تشير فترة الثقة الى المدى الذي يقترب منه (الاستدلال به) المعلمة المقدرة من معلمة المجتمع الاحصائي (المعلمة الحقيقية)

$$Pr\left[\hat{a} - t_{(\frac{\alpha}{2}, n-k)} * S(\hat{a}) \leq a \leq \hat{a} + t_{(\frac{\alpha}{2}, n-k)} * S(\hat{a})\right]$$

$$Pr\left[\hat{b} - t_{(\frac{\alpha}{2}, n-k)} * S(\hat{b}) \leq b \leq \hat{b} + t_{(\frac{\alpha}{2}, n-k)} * S(\hat{b})\right]$$

اي احتمال ان تقع القيمة الحقيقية داخل حدود الثقة 95%

ان حدي الثقة تشير الى تقدير فترة بحيث ان من كل (100) محاولة تقع القيمة الحقيقية بين الحدين الاصغر والاكبر سواء (95) مرة او (99) مرة اذا كان مستوى المعنوية وتبتعد القيمة الحقيقية عن هذين الحدين (5 مرات) او (مرة واحدة)

مستوى الثقة هو احتمال ان تقع قيمة المعلمة الحقيقية داخل فترة الثقة

$Pr = 1 - \alpha$

(38) F test and Analysis of variance

(6) اختبار F وجدول تحليل التباين

سمي باختبار F نسبة الى فيشر (مكتشفه) ويستخدم اختبار F^* الى جدول تحليل التباين ANOVA table

Analysis of variance

ووفقاً لهذا الجدول يتم تحليل التباين الكلي الى مركبتين رئيسيتين هما RSS و ESS وكما التباين المفسر موضحة في الجدول التالي

مصادر التباين Source of variation S.V	مجموع المربعات Sum of squares S.S	درجات الحرية Degrees of freedom (d.f)	متوسط مربع الخطأ mean square error M.S.M	F^*
Explained variations	$\sum\hat{y}^2 = RSS = \hat{b}\sum xy = \hat{b}^2\sum x_i^2 = \sum(\hat{Y}-\bar{Y})^2$	$K-1$	$\frac{RSS}{K-1}$	$F^* = \frac{RSS/K-1}{ESS/n-K}$
unExplained variations	$\sum e^2 = ESS = \sum y^2 - \sum\hat{y}^2 = \sum(Y-\hat{Y})^2$	$n-K$	$\frac{SSE}{n-K}$	
Total variation	$\sum y^2 = \sum\hat{y}^2 + \sum e_i^2 = \sum Y^2 - n\bar{Y}^2 = \sum(Y-\bar{Y})^2$	$n-1$		

يستخدم اختبار F^* الى توزيع F وتوزيع (F) هو توزيع مشروط ... من صفر و α فعند تقدير قيمة F تقارن قيمة F^* المحسوبة مع قيمتها الجدولية وعند مستوى معنوية α ودرجة حرية مقدارها $(K-1)$ للبسط و $(n-K)$ للمقام

$$F^* > F_{\text{الجدولية}}$$

نقبل بالفرض البديل ونرفض العدم مما يشير الى معنوية العلاقة الدالية ككل

اما عندما $F^* < F_{\text{الجدولية}}$

نقبل الفرض العدم ونرفض البديل $H_0: B=0$ مما يشير الى عدم معنوية العلاقة الدالية ككل.

٧- تقدير تباين حد الخطأ العشوائي

Estimated variance of Random Error

لاجراء اختبار معنوية المعالم المقدرة يتم استخدام تباين a و b والان تلك التباينات تحتوي على معلمة تباين المجتمع σ^2 وهي غير معروفة، وللتقدير ذلك يتم استخدام تباين حد الخطأ العشوائي والذي يرمز له بـ S_e^2 وذلك لان σ^2 تمثل تباين المجتمع وان المقدر الحقيقي لها هو S_e^2

$$\hat{\sigma}_u^2 = \frac{\sum e_i^2}{n-K} = \frac{\sum y^2 - \hat{b}\sum xy}{n-K}$$

$$= \frac{\sum y^2 - \hat{b}^2 \sum x_i^2}{n-K}$$

$$\sum e^2 = \sum y^2 - \hat{b}\sum xy \quad \text{or} \quad \sum e^2 = \sum y^2 - \hat{b}^2 \sum x_i^2$$

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

$$= Y_i - \hat{a} - \hat{b}X_i \qquad \text{بالتعويض عن } \hat{Y}_i$$

$$= Y_i - \bar{Y} - \hat{b}\bar{X} - \hat{b}X_i \qquad \text{بالتعويض عن } \hat{a}$$

$$e_i = y_i - \hat{b}x_i$$

$$\sum e_i^2 = \sum (y_i - \hat{b}x_i)^2$$

$$= \sum y_i^2 - 2\hat{b}\sum y_i x_i + \hat{b}^2 \sum x_i^2$$

نضرب ونقسم الحد الاوسط بـ $\frac{\sum x_i^2}{\sum x_i^2}$

$$= \sum y_i^2 - 2\hat{b}\frac{\sum y_i x_i \sum x_i^2}{\sum x_i^2} + \hat{b}^2 \sum x_i^2$$

$$= \sum y_i^2 - 2\hat{b}^2 \sum x_i^2 + \hat{b}^2 \sum x_i^2$$

حيث ان $\hat{b} = \frac{\sum y x_i}{\sum x_i^2}$

$$\boxed{\sum e_i^2 = \sum y^2 - \hat{b}^2 \sum x_i^2}$$

$$\boxed{\hat{\sigma}_u^2 = \frac{\sum y^2 - \hat{b}^2 \sum x_i^2}{n-K}}$$

او بالتعويض عن قيمة $\hat{b}$ في الحد الاخير

$$\hat{b} = \frac{\sum y x_i}{\sum x_i^2}$$

$$\sum e_i^2 = \sum y_i^2 - 2\hat{b}\sum y x_i + \hat{b}\frac{\sum y x_i}{\sum x_i^2}\sum x_i^2$$

$$\sum e_i^2 = \sum y_i^2 - 2\hat{b}\sum y x_i + \hat{b}\sum y x_i$$

$$\boxed{\sum e_i^2 = \sum y_i^2 - \hat{b}\sum y x_i}$$

$$\boxed{\hat{\sigma}_u^2 = \frac{\sum y^2 - \hat{b}\sum y x_i}{n-K}}$$