خصائص مقدرات المربعات الصغرى

خاصية عدم التحيز

التحيز هو ذلك الانحراف بين مقدرة ما ووسط توزيعها، فإذا كان هذا الفرق يختلف عن الصفر نقول عن ذلك المقدر بأنه متحيز. وإذا عدنا إلى مقدرتي المربعات الصغرى فإننا نجد الفرق معدوما ومنه نقول أنهما مقدرتين غير متحيزتين ولاعتبارت رياضية نقبل بدون برهان أنهما غير متحيزين (لمن أراد الاطلاع على البرهان يمكنالرجوع لكتاب الأستاذ شيخي محمد)

خاصية الاتساق

تنطلق هذه الفكرة من نظرية Gauss-Markov والتي تقول" من بين المقدرات الخطية وغير المتحيزة، تكون مقدرتا المربعات الصغرى العادية $\hat{\beta}_0$ و $\hat{\beta}_1$ أفضل مقدرتين خطيتين اذا كان لها أصغر تباين ممكن مقارنة مع بقية المقدرات الخطية غير المتحيزة الأخرى.

5.2.معامل التحديد R^2 أو القوة التفسيرية للنموذج(تحليل التباين أنوفا **Anova**):

تساعد البواقي e_i على قياس قدرة تمثيل المعادلة في النموذج لمشاهدات العينة، حيث أن كلما كبرت قيمة البواقي كلما دل ذلك على ضعف التمثيل والعكس صحيح، إن المشكلة في استعمال البواقي كمقياس لجودة التوفيق هو أن قيمة البواقي تعتمد على المتغير التابع Y_i، الذي نعرفه حول وسطه انطلاقا من الشكل رقم (1) كما يلي :

$$Y_i = \hat{Y}_i + ei$$
$$Y_i - \overline{Y} = \hat{Y}_i - \overline{Y} + ei$$

وبتربيع طرفي المعادلة السابقة وإدخال المجموع عليها نتحصل على ما يلي:

$$\sum_i (Y_i - \overline{Y})^2 = \sum_i (\hat{Y}_i - \overline{Y})^2 + \sum_i e_i^2$$

كيفية الحصول على المعادلة الأخير مبين في البرهان الموالي

$$y_i - \overline{y} = \hat{y} - \overline{y} + e_i$$

$$\Sigma(y_i - \overline{y})^2 = \Sigma(\hat{y} - \overline{y} + e_i)^2$$

$$\Sigma(y_i - \overline{y})^2 = \Sigma\left[(\hat{y} - \overline{y})^2 + e_i{}^2 + 2(\hat{y} - \overline{y})(e_i)\right]$$

$$\Sigma(y_i - \overline{y})^2 = \Sigma\left[\hat{y} - \overline{y})^2 + e_i{}^2 + 2(\hat{y} - \overline{y})(e_i)\right]$$

$$= \Sigma(\hat{y} - \overline{y})^2 + \Sigma e_i{}^2 + 2\Sigma(\hat{y} - \overline{y})(e_i)$$

$$= \Sigma(\hat{y} - \overline{y})^2 + \Sigma e_i{}^2 + 2\Sigma e_i(\hat{y} - \overline{y})$$

$$\boxed{\underbrace{\Sigma(y_i - \overline{y})^2}_{SCT} = \underbrace{\Sigma(\hat{y} - \overline{y})^2}_{SCE} + \underbrace{\Sigma e_i{}^2}_{SCR}}$$

SCT	=	SCE	+	SCR

بقسمة طرفي المعادلة الأخيرةعلى SCT:

نجد:

$$\boxed{\begin{aligned} 1 &= \frac{SCE}{SCT} + \frac{SCR}{SCT} \\ 1 &= \frac{\Sigma(\hat{y} - \overline{y})^2}{\Sigma(y_i - \overline{y})^2} + \frac{\Sigma \ell i^2}{\Sigma(y_i - \overline{y})^2} \end{aligned}} \quad (\star)$$

المقدار $\frac{SCE}{SCT}$ أو $\frac{\Sigma(\hat{y}-\bar{y})^2}{\Sigma(y_i-\bar{y})^2}$ يعرف "بمعامل التحديد" R^2 ومنه (✯) تصبح كمايلي:

$$\boxed{1 = R^2 + \frac{\Sigma e_i^2}{SCT}} \Rightarrow \boxed{R^2 = 1 - \frac{\Sigma e_i^2}{SCT}} \quad \text{—}(✯_1)$$

ملاحظات:

من العلاقة ($✯_1$) يمكن أن نسجل جملة الملاحظات التالية:

- R^2 هو نسبة موجبة (وبالتالي هو رقم موجب) محصور بين 0 و 1 $(0 < R^2 < 1)$.

كلما كان معامل التحديد (R^2) قريبا من (1) كلما كان النموذج قويا والعكس صحيح.

من العلاقة السابقة نستنتج أيضا أنه كلما كانت النسبة $\frac{SCR}{SCT}$ [تساوي $= \frac{\Sigma ei^2}{\Sigma(y_1-\bar{y})^2}$] صغيرة كلما كان R^2 قويا لأنه في هذه الحالة سيقترب من (1) الصحيح والعكس صحيح.

6.2. معامل الارتباط: (R أو r)

كما أن معامل التحديد R^2 هو مؤشر يقيس لنا القوة التفسيرية للنموذج المقدر $\hat{y}$ أو بالأحرى يقيس جودة النموذج. فإن معامل الارتباط R أو r هو كذلك مؤشر أو مقياس يقيس لنا درجة الارتباط أو قوة العلاقة بين المتغير التابع (y) والمتغير المستقل (x).

معامل الارتباط هو الجذر التربيعي لمعامل التحديد: $\boxed{r = \sqrt{R^2}}$

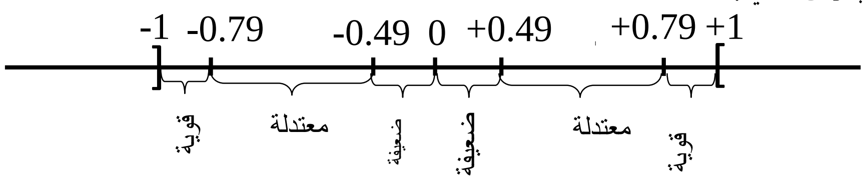
كذلك معامل الارتباط (R) بالإضافة إلى قياسه لقوة العلاقة بين x و y فهو أيضا يبين لنا اتجاه أو نوع العلاقة بينهما أي يبين إشارة او اتجاه العلاقة (+ / −) وهو كذلك عبارة عن نسبة محصورة بين طرفي المجال المفتوح

]-1 ، 1[

لتبيان إتجاه العلاقة بين x و y يتم حساب معامل الارتباط (R) بالاعتماد على العلاقة التالية:

$$\boxed{R = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{\sqrt{(n\Sigma x_i^{2} - (\Sigma x_i)^2)(n\Sigma_i)^2)}}}$$

حيث تعطى المجالات التالية:



7.2.دراسة معنوية معامل الارتباط:

الهدف من دراسة معنوية معامل الارتباط هو الحكم على معامل هذا الارتباط فيما إذا كان معنويا أو غير معنوي (لا معنوي) ومعناه هل هو مقبول إحصائيا (دال) أو نقول معنوي

أو غير مقبول إحصائيا (غير دال) = لا معنوي

للقيام بهذا الاختبار: أولا: نفرض أن معامل الارتباط الحقيقي هو الرمز (ℓ) [وهو معامل مجهول ويدعى معامل بيرسون Pearson]

ثانيا: يلزمنا فرضيتين "H_0": الفرضية الابتدائية (الأولية) و

"H_1": الفرضية البديلة.

حيث نضع الجملة التالية:

$$\begin{cases} H_0 : \ \ell = 0 & \text{(الارتباط غير موجود أو لا معنوي)} \\ H_1 : \ \ell \neq 0 & \text{(الارتبـاط موجـود أو معنوي)} \end{cases}$$

Pearson correlation significance test

$$\begin{cases} H_0 : \varphi \ = 0 \ \ \ldots.. \ (no\ significance) \\ H_1 : \varphi \neq 0 \ \ \ldots\ldots \ \ (significante\ ) \end{cases}$$

نحتاج إلى ما يعرف بدالة الاختبار وتعرف ب tالمحسوب رمزها "t_c" والى t الجدولي أو النظري رمزه"t_t"

حيــث:

$$t_t = t_v^{1-\frac{\alpha}{2}} = t_{n-k}^{1-\frac{\alpha}{2}}$$

$$V=n-k$$

K: عدد المعالم المقدرة وهما معلمتين فقط كما بين ذلك سابقا ، لذلك في نموذج الانحدار البسيط دائما يكون :

k=2

أما t_c فيحسب بقيمته المطلقة كما يلي :

$$t_c = r\sqrt{\frac{n-2}{1-r^2}}$$

إذا وجدنا : $|t_c| \geq t_t$ في هذه الحالة نرفض H_0 لصالح H_1 عند مستوى معنوية α% حيث مستوى المعنوية عادة ما يأخذ القيم المشهورة التالية:(α= 1%, 2.5% , 5%, 10%)

* أما إذا حدث العكس أي $|t_c|$ أصغر تماما من t_t: فإننا نقبل H_0 ونرفض H_1 أي الارتباط لا معنوي، أو نقول غير مقبول عند مستوى معنوي α%.

<u>مثال</u>