برهان أساس للفترة [-1,1] لمعامل الارتباط

ا. سليم اسماعيل الغرابي
كلية الادارة والاقتصاد
جامعة بغداد

1- **مقدمة**:

مفهوم معامل الارتباط كمقياس يربط بين متغيرين هذا يجلب انتباهنا إلى موضوع الإحصاء في كل المستويات. أكثر من ذلك هناك ثلاث نقاط خاصة هي اعتيادياً نشدد عليها كما يأتي: -

1) معامل الارتباط هو الدليل المعياري والذي قيمته لا تعتمد على قياسات المتغيرات الأصلية.

2) قيمته تقع في المدى[1,1-] .

3) أن مربع قيمته نصف نسبة تقليل في أحد المتغيرات بينما الآخر يبقى ثابتاً.

ولو أن هذه الخواص هي على نحو مستقيم ولو أن فئات رياضية من الطلاب [أنظر: Mendenhall et.al.1981, chap-5]هم غالباً يعتقدون كحقيقة في مقدمة كثير من المواضيع خصوصاً تلك التي توجه الى الحقول التي أحصائياً هي التي تطبق.

نسبة إلى تلك الخلفية، الغرض في هذا البند هو اشتراط أساسي لبرهان أن معامل الارتباط يقع في الفقرة [1,1-] بواسطة العمل مع تباينات جمع وفرق متغيرين عشوائين معيارين.

هكذا مناقشة تشترط طريقة توصل الخواص المشتركة لمعامل الارتباط لغرض المستمعين من الطلاب.

بعض الدلالة لأجل انحدار خطي يتم إحضاره كوصف قانون معامل الارتباط لمقياس لإنقاص نسبة أمكانية التحويل.

2) **البرهان**:

أفرض أن (x_1,y_1) , (x_2,y_2)......, (x_n,y_n) عينة من أزواج المشاهدات من الموضوعات:

أعتبر العلاقة: $u_i = ax_i + cy_i$ التي هي دوال خطية لـ x_i، y_i لمواضيع خاصة.

معدل العينة لـ $\{ u_i \}$ هو

$$\bar{u} = \frac{1}{n}\sum_{i=1}^{n}(ax_i + cy_i) = a\bar{x} + c\bar{y} \qquad ....(1)$$

تباين العينة لـ $[u_i]$ هو

$$S_u^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}[(a \times i + cyi) - (a\bar{\times} + c\bar{y})]^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}[a(x_i - \bar{x}) + c(yi - \bar{y})]^2$$

$$= a^2 S_x^2 + 2ac\, S_{xy} + c^2\, S_y^2 \qquad ....(2)$$

حيث أن

$$S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

، $$S_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

، $$S_{\times y} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

الصيغة لأجل S_u^zترينا القانون للتباين المشترك لـ S_{xy} في التباين لـ $\{ u_i \}$

الآن: أفرض أن

$$a = \frac{1}{S_x}, \qquad C = \frac{1}{S_y}$$

فأن

$$S_u^2 = 2(1 + r_{xy})$$

حيث أن

$$r_{xy} = \frac{S_{\times y}}{S_x S_y}$$

هو معامل الارتباط للعينة وبالمثل إذا كانت $a = \frac{1}{S_x}, C = -\frac{1}{S_y}$

فأن

$$S_u^2 = 2(1-r_{xy})$$

حيث أن $S_u^2 \geq o$ لكل الاختيارات لـ (a,c)

$$(1-r_{xy}) \geq 0 \text{ ، } (1-r_{xy}) \geq o$$

وعليه فأن $-1 \leq r_{xy} \leq 1$ عندما

$$a = \frac{1}{S_x} \quad ,C = \frac{1}{S_y}$$

حول المتغيرات $f_i = ax_i \,, g_i = cy_i$

هي معيارية لتكون $S_f^2 = S_g^2 = 1$

أنه يتبع أن $S_{fg} = r_{fg}$

لأجل هذه الحالة.

أكثر تعميم، تعريف معامل الارتباط يؤدي إلى أن $r_{xy} = r_{fg}$

لأي اختيار لـ (a,c) وهكذا r_{xy} هو الأساس لتمثيل التباين المشترك للمتغيرات العشوائية الثابتة لتكاثر التغيرات في القياس.

من المناقشة السابقة، أنه يمكن ملاحظة أن $r_{xy} = 1$

هذا يؤدي إلى أن $S_u^2 = o$

لأجل $a = \frac{1}{S_x}, C = -\frac{1}{S_y}$ وهذا يعني أن:

$$\sum_{i=1}^{n} \left\{\frac{(x_i-\overline{x})}{S_x} - \frac{(y_i-\overline{y})}{S_y}\right\}^2 = o \qquad ....(3)$$

ذلك أن $\frac{(x_i-\bar{x})}{S_x} = \frac{(y_i-\bar{y})}{S_y}$

$$\text{or } y_i = \bar{y} + \frac{S_y}{S_x}(x_i-\bar{x}) \qquad ....(4)$$

هذه النتيجة تعني أن $r_{xy}=1$

نسبة إلى الحالة التي هي أن { y_i } هي دالة خطية متزايدة لـ { x_i }،عندما $r_{xy}= -1$

هذه بعض النقاط تطبق بالنسبة للدوال الخطية المتناقصة

الآن أفرض أن $c=1$

فأن

$$S_u^2 = a^2 S_x^2 + 2aS_{xy} + S_y^2 \qquad via(2)$$

ما هي القيمة لـ a التي تجعل S_u^2 أقل ما يمكن .

هنا السؤال يمكن يعنون بواسطة كتابة :

$$S_u^2 = a^2 S_x^2 + 2aS_{xy} + (\frac{S \times y}{S_x})^2 - (\frac{S_{xy}}{S_x})^2 + S_y^2$$

$$= (aS_\times + \frac{S_{\times y}}{S_x})^2 + S_y^2 - (\frac{S_{\times y}}{S_x})^2 \qquad ....(5)$$

وهكذا S_u^2 هي أقل ما يمكن (نهاية صغرى) عندما

$$[aS \times + (S_{xy} / S_x)] = o$$

أو

$$a = (-S_{xy} / S_x^2)$$

القيمة الصغرى في هذه الحالة هي :

$$S_u^2 = S_y^2 - (S_{xy}^2 / S_x^2) = S_y^2 (1 - r_{xy}^2) \qquad ....(6)$$

بما أن S_u^2 هي نهاية صغرى بواسطة

$$a = (-S_{xy} / S_x^2)$$

هي مركبة الحدود

$$[a(x_i-\bar{x}) + (y_i-\bar{y})]^2$$

يمكن أن ينظر لها نهاية صغرى على كل القيم

وعليه

$$[y_i - \bar{y} + a(x_i - \bar{x})]^2$$

هي نهاية صغرى على كل القيم هذه الخاصية يمكن إن تفسر مؤشر ذلك أن y_i هي مخبـر جيد بواسطة:

$$\hat{y}_i = \bar{y} + (S_{xy} / S_x^2)(x_i - \bar{x}) \qquad ....(7)$$

فأن

$$S_u^2 = S_y^2 (1 - r_{xy}^2)$$

قياسات تبقى قابلة للتغيير لمتتابعة جزئية لهذا مستند في التفسير بشرط استعمال هكذا تأكيد وأن .

$$(S_y^2 - S_u^2) / S_y^2 = 1 - (S_u^2 / S_y^2) = r_{xy}^2 \qquad ....(8)$$

القياسات النسبية تصغر في أمكانية التحويل من خلال استعمال هكذا تنبؤ.

نفس المناقشة تستخدم في هذا المبحث لترينا أن عينة الارتباط r_{xy} تقع في الفقـرة [1,1-]. يمكن أيضاً تطبيق معامل الارتباط ρ في المجتمع لأجل التوزيعات المزدوجـة (المـشتركة) لمتغيري (X,Y). التبديل الرئيس هو تبديل المجتمع المفرد في (2). بواسطة توقـع قيمـة مؤشره في حالة التوزيعات غير المستمرة هذه يمكن ببساطة تجمع خلال النتـائج الممكنـة للتوزيعـات المـشتركة. ملاحظـات إضـافية فـي هـذا المجـال معطـاة فـي [Freund (1971)p.368,exercise 11] .

Reference

1) Freund J. E. (1971) Mathematical Statistics (2 nd. Ed.) Englewood Cliffs. NJ Prentice – Hall.
2) Mendenhall W., Scheaffer, R. L. and Wackerly, D. D. (1981) Mathematical Statistics With Applicieations (2nd, ed.) Boston, Daxbury press.

3) المشهداني، محمود حسن، هرمز، امير حنا (1989) الاحصاء – مطبعة التعليم العالي في الموصل.