مبادئ الاحصاء والاحتمالات

Principle of Statistic and Probability

MS 304

محاضرة رقم 6

الإرتباط Correlation والإنحدار Regression

معامل الارتباط الخطي لبيرسون

يستخدم معامل الارتباط الخطي لبيرسون لقياس التغير الذي يطرأ على المتغير Y عندما تتغير قيم X و العكس . في حالة البيانات المباشرة ، إذا كان لدينا زوج المشاهدات الآتية من مجتمع : $(x_1,y_1),(x_2,y_2),\ldots,(x_3,y_3)$ فإن معامل الارتباط لبيرسون يرمز له r و يعطى من خلال العلاقة الآتية

$$r = \frac{1}{n} \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} \quad ...(1)$$

σ_x , σ_y هما الانحراف المعياري للمتغيرين y و x على الترتيب . في حالة البيانات المأخوذة من عينة فإن العلاقة السابقة تصبح كالآتي

$$r = \frac{1}{n-1} \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} \quad ...(2)$$

لاحظ : معامل الارتباط يساوي صفر عندما تكون الظاهرتان مستقلتين و يكون قوياً عندما يقترب معامل الارتباط من الواحد الصحيح و ضعيفاً عندما يقترب من الصفر .لإيجاد معامل ارتباط بيرسون حسابياً يفضل استخدام الصيغة الآتية

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{\left(n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right)\left(n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right)}} \quad ...(3)$$

مثال 1

الجدول الآتي يبين درجات مجموعة مكونة من ثمانية طلاب في كلٍ من مادتي الإحصاء X و الرياضيات Y . هل توجد علاقة بين تحصيل الطلاب في المادتين ؟

X	Y
13	15
9	7
19	17
15	15
11	10
8	9
16	14
11	10

الحل: لتبسيط البيانات التي بالجدول نطرح مقداراً ثابتاً $a = 10$ من كل من قيم

X و Y و من ثم نكون الجدول الآتي

X	Y	$x' = x - 10.$	$y' = y - 10.$	$x'y'$	x'^2	y'^2
13	15	3	5	15	9	25
9	7	1-	3-	3	1	9
19	17	9	7	63	81	49
15	15	5	5	25	25	25
11	10	1	0	0	1	0
8	9	2-	1-	2	4	1
16	14	6	4	24	36	16
11	10	1	0	0	1	0

نعوض عن قيم $x = x'$ و $y = y'$ في معادلة 3 لنحصل على معامل ارتباط بيرسون

$$r = \frac{8\times132 - 22\times17}{\sqrt{\left(8\times158-(22)^2\right)\left(8\times125-(17)^2\right)}}$$

$$r = \frac{690}{744.7} = 0.93 \approx 1$$

و هذا يعني وجود ارتباط قوي جداً بين درجات تحصيل الطلاب في المادتين .

معامل الارتباط للرتب لسبيرمان Spearman

معامل الارتباط الخطي لبيرسون الذي سبق الحديث عنه يقيس مقدار قوة الارتباط بين متغيرين وذلك في حالة البيانات الكمية فقط . في بعض الأحيان يكون مطلوباً إيجاد قوة الارتباط بين متغيرين على صورة بيانات وصفية مثل تقديرات الطلاب في مادتين مختلفتين معطاة بالأحرف A,B,C,D,E فإنه يصعب حساب معامل الارتباط بطريقة بيرسون ، لذلك نشأت الحاجة إلى إيجاد مقياس يعطي قوة الارتباط للبيانات الوصفية وهذا المقياس هو ما يسمى بمعامل

ارتباط الرتب لسبيرمان و هو يعطي مقياس للارتباط في كل من البيانات الكمية و الوصفية التي لها صفة الترتيب مثل تقديرات الطلاب فإنه يمكن إعطاءها رتب من حيث كبر التقدير أو صغره . نلاحظ أن رتب المتغيرين (X, Y) تزيد و تنقص حسب زيادة و نقص قيمهما ، لذلك فإن حساب معامل الارتباط للرتب يقترب كثيراً من معامل الارتباط لبيرسون ولكن يمتاز عنه في السهولة و الدقة و خاصة عندما تكون أزواج القيم أقل من 30 و يعطى معامل ارتباط الرتب بالعلاقة الآتية

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

حيث r_s معامل ارتباط الرتب لسبيرمان لعدد n من أزواج القيم (X,Y) و

$d_i = a_i - b_i$ حيث a_i رتب المتغير X و b_i رتب المتغير Y . ويمكن

توضيح حساب الرتب كالآتي : المقصود بالرتب هنا هو إيجاد رتب القراءات

(X,Y) مع بقاء كل قراءة مكانها و ذلك بأن نتصور ترتيب البيانات تصاعدياً

أو تنازلياً ، ففي حالة الترتيب التصاعدي تأخذ رتبة أصغر قراءة الرقم 1 في

حين تأخذ القراءة التي تليها الرقم 2 وهكذا . في حالة تساوي قيمتان نعطي لكل

قيمة منهما رتبة تساوي الوسط الحسابي لرتبتيهما .

مثال 2

أوجد رتب X التي قيمها معطاة في الجدول الآتي

X	10	4	5	7	2

الحل

نتصور ترتيب قيم X تصاعدياً كما في الجدول الآتي

X	10	4	5	7	2
الرتبة	2	9	15	11	2

مثال 3

أوجد رتب التقديرات الآتية

$$B, C, B, E, D, D, A$$

الحل

إذا تصورنا E تأخذ الرتبة 1 ، D مكررة و من ثم تأخذ الرتبتين 2 و 3 و عليه

كل قيمة للتقدير D تأخذ الرتبة $\frac{2+3}{2}=2.5$ ، C تأخذ الرتبة 4 ، B مكررة و

تأخذ الرتبة $\frac{5+6}{2}=5.5$ وأخيراً A تأخذ الرتبة 7 كما في الجدول الآتي

X	A	D	D	E	B	C	B
الرتبة	7	2.5	2.5	1	5.5	4	5.5

مثال 4

أوجد معامل ارتباط الرتب لدرجات الطلاب في مادتي الإحصاء و الرياضيات:

إحصاء	13	9	19	15	8	16	11
رياضيات	15	7	17	10	9	14	10

الحل

نكون الجدول الآتي

إحصاء	رياضيات	a	b	d	d^2
13	15	5	6.5	–1.5	2.25
9	7	2	1	1	1
19	17	8	8	0	0
15	15	6	6.5	–0.5	2.25

11		3.5	3.5	0	0
8	9	1	2	–0.5	1
16	14	7	5	2	4
11	10	3.5	3.5	0	0
$\sum$					8.5

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6\times 8.5}{8(64-1)} = 1 - \frac{51}{504} = 0.9$$