الباب الخامس
العلاقات الإحصائية (الارتباط والانحدار)

مقدمة

تناولنا فى الأبواب السابقة طرق دراسة متغير واحد لأى ظاهرة محل الدراسة، مثل أوزان مجموعة من الطلاب أو أجور مجموعة من العمال... إلخ. وعرضنا كيف يمكن تلخيص البيانات فى جداول توزيعات تكرارية وكيفية عرضها بيانيا. كذلك دراسة بعض المقاييس العددية التى تساعد على معرفة بعض خصائص التوزيعات التكرارية، ومنها مقاييس النزعة المركزية ومقاييس التشتت والالتواء والتفرطح.

سوف نتناول الآن دراسة البيانات التى يكون لأفرادها متغيران يتغيران معا فى وقت واحد، وذلك لمعرفة نوع العلاقة التى تربط بينهما، مثل دراسة العلاقة بين أوزان وأطوال مجموعة من الطلاب أو أعمار ودرجات مجموعة من الطلاب، وهكذا.... ثم إيجاد مقاييس تقيس درجة هذه العلاقة.

كذلك سوف نقوم بدراسة العلاقة بين المتغيرين (x,y)، فإذا كانت هناك علاقة بين المتغير x والمتغير y فكيف يمكن التعبير عنها بمعادلة رياضية ومنها يمكن التنبؤ بقيمة أحد المتغيرين إذا علمت قيمة المتغير الآخر. وسوف نتناول فى هذا الباب إيجاد مقاييس لقياس قوة الارتباط بين المتغيرين (x,y) فى الحالة الخطية فقط. وسندرس منها معامل الارتباط الخطى لبيرسون (Pearson)، ومعامل ارتباط الرتب لسبيرمان (Spearman)، كما سوف ندرس معامل الاقتران ومعامل التوافق، وكذلك دراسة معادلة الانحدار الخطى البسيط.

٥-١ معامل الارتباط الخطى لبيرسون

يستخدم معامل الارتباط الخطى لبيرسون لقياس التغير الذى يطرأ على المتغير y عندما تتغير قيم x أو العكس. ويستخدم عادة فى حالة البيانات الكمية.

إذا كان لدينا أزواج المشاهدات التالية:

$$(x_1,y_1)\ ,\ (x_2,y_2)\ ,\ \ldots,(x_n,y_n)$$

فإن معامل الارتباط r لبيرسون يعطى من خلال العلاقة:

$$r = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sqrt{(\sum x^2 - \frac{(\sum x)^2}{n})(\sum y^2 - \frac{(\sum y)^2}{n})}}, \qquad (1)$$

ويكون لمعامل الارتباط (r) الخصائص التالية:

١- قيمته تساوى صفرا عندما تكون الظاهرتان مستقلتان تماما.

٢- قيمته مقدار موجب عندما يكون الارتباط بين المتغيرين طرديا. ويكون قويا عندما يكون المقـدار الموجب قريبا من الواحد الصحيح، وضعيفا عندما يكون المقدار الموجب قريبا من الصفر.

٣- قيمته مقدار سالب عندما يكون الارتباط بين المتغيرين عكسيا. ويكون قويا عندما يكون المقـدار السالب قريبا من (-١)، وضعيفًا عندما يكون المقدار السالب قريبا من الصفر.

مثال (١): الجدول التالى يوضح درجات مجموعة مكونة من ثمانية طلاب فى كل من مادتى الإحصاء والرياضيات فى إحدى الامتحانات. هل هناك علاقة بين تحصيل الطالب فى المادتين؟

الإحصاء x	13	9	19	15	11	8	16	11
الرياضيات y	15	7	17	15	10	9	14	10

الحل: لتبسيط البيانات بالجدول نطرح مقدار ثابت = ١٠ من كل من قيم x وقيم y ونكون الجـدول التالي:

x	y	$-10\, x = x$	$-10\, y = y$	$y\ x$	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
		22	17	132	158	125

$$r = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sqrt{(\sum x^2 - \frac{(\sum x)^2}{n})(\sum y^2 - \frac{(\sum y)^2}{n})}}$$

$$r = \frac{132 - \frac{(22 \times 17)}{8}}{\sqrt{(158 - \frac{(22)^2}{8})(125 - \frac{(17)^2}{8})}}$$

$$r = 0.93$$

أى يوجد ارتباط طردى (قوى جدا) بين درجات تحصيل الطالب فى المادتين.

٢-٥ معامل ارتباط الرتب (سبيرمان)

معامل الارتباط الخطى لبيرسون الذى سبق الحديث عنه يقيس مقدار قوة الارتباط بين متغيرين وذلك فى حالة البيانات الكمية. لكن فى بعض الأحيان يكون مطلوب إيجاد قوة الارتباط بين متغيرين على صورة بيانات وصفية يمكن وضعها فى صورة ترتيبية، مثال على هذا تقديرات الطلاب فى مادتين مختلفتين، فيكون من الصعب حساب معامل ارتباط بيرسون. لذلك نشأت الحاجة إلى إيجاد مقياس يعطى قوة الارتباط للبيانات الوصفية. وهذا المقياس هو ما يسمى بمعامل ارتباط الرتب لسبيرمان، وهو يعطى مقياسا للارتباط فى كل من البيانات الكمية والوصفية التى لها صفة الترتيب مثل تقديرات الطلاب، فإنه يمكن إعطاء رتب لها من حيث كبر التقدير وصغره وكذلك البيانات الكمية. نلاحظ أن رتب المتغيرين (x,y) تزيد وتنقص حسب زيادة ونقص كل من قيم المتغيرين (x,y). لذلك فإن حساب معامل الارتباط للرتب يقترب كثيرا من معامل ارتباط بيرسون، ولكن يمتاز عنه فى السهوله والدقة خاصة عندما تكون أزواج القيم أقل من ١٥.

ويعطى معامل ارتباط الرتب بالعلاقة التالية:

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

حيث r_s معامل ارتباط الرتب لسبيرمان، n تمثل عدد أزواج القيم (x,y)، d هى الفرق بين رتب أزواج القيم (x,y) ويمكن توضيح ذلك من خلال المثال التالي:

مثال (٢): اوجد معامل ارتباط الرتب لتقديرات الطلاب فى كل من مادتى الإحصاء والرياضيات كما هو موضح بالجدول التالي:

الرياضيات x	A	C	C	C	B	D
الإحصاء y	B	B	D	C	A	E

الحل: نقوم بتلخيص الحل فى الجدول التالي:

الرياضيات x	الإحصاء y	رتبة x $= a$	رتبة y $= b$	$d = a - b$	d^2
A	B	6	4.5	1.5	2.25
C	B	3	4.5	-1.5	2.25
C	D	3	2	1	1
C	C	3	3	0	0
B	A	5	6	-1	1
D	E	1	1	0	0
					6.5

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$
$$r_s = 1 - \frac{6 \times 6.5}{6(36-1)}$$
$$r_s = 1 - 0.186$$
$$r_s = 0.814$$

أى يوجد ارتباط طردى قوى بين تقديرات مادتى الرياضيات والإحصاء.

٣-٥ معامل الاقتران ومعامل التوافق

لقد سبق أن وضحنا بأن معامل ارتباط بيرسون يعطى قوة الارتباط فى حالة البيانـات الكميـة، وكذلك معامل ارتباط الرتب لسبيرمان، وهو يستخدم لإيجاد قوة الارتباط للرتب فـى حالـة البيانـات الكمية والوصفية التى لها صفة الترتيب، ولكن قد تكون هناك بيانات وصفية لها صفات مميزة ولكن لا يمكن ترتيبها مثل الحالة الاجتماعية، وكذلك لون البشرة....إلخ. ولقياس قوة الارتباط لهـذه البيانـات نشأت الحاجة إلى إيجاد مقياس مناسب يقيس الارتباط بين هذه الصفات، ومنها معامل الاقتران الـذى يستخدم عندما يكون لكل من الظاهرتين صفتين فقط. كذلك دراسة معامل التوافق فى حالة تكون كـل من الظاهرتين أو إحداهما على الأقل من أكثر من صفتين كما سنوضح ذلك بالتفصيل كما يلي:

١- معامل الإقتران

يستخدم معامل الاقتران لقياس قوة الارتباط بين ظاهرتين كل منهما ذات صفتين فقط، وسـوف يرمز له بالرمز C.C مثل دراسة قوة الارتباط بين التدخين والتعليم، حيث يوضـح الجـدول التـالى التكرار للصفات:

التدخين / التعليم	يدخن	لا يدخن
متعلم	A	B
غير متعلم	C	D

فيكون معامل الإقتران C.C كالتالي:

$$C.C = \frac{AD - BC}{AD + BC}, \qquad (3)$$

ونوضح ذلك بالمثال الآتي:

مثال (٣): عند دراسة العلاقة بين التعليم والتدخين فى إحدى الشركات أخذت عينة مكونة من ١٧ شخصا وكانت النتائج على النحو التالي:

التدخين / التعليم	يدخن	لا يدخن
متعلم	5	5
غير متعلم	3	4

احسب معامل الاقتران بين التدخين والتعليم.

الحل:

$$C.C = \frac{AD - BC}{AD + BC}$$

$$C.C = \frac{(5\times 4) - (3\times 5)}{(5\times 4) + (3\times 5)}$$

$C.C = 0.14$ وهو ارتباط ضعيف

٢- معامل التوافق

إذا كانت بيانات الظاهرتين التى لدينا عبارة عن بيانات وصفية لكل منهما أو إحداهما على الأقل وكانت مقسمة لأكثر من صفتين، فإن معامل الاقتران السابق لا يصلح فى هذه الحالة ويتم استخدام مقياس آخر هو معامل التوافق C. لحساب معامل التوافق نفرض أن لدينا الظاهرة X والتى لها r من الصفات والظاهرة الثانية Y والتى لها s من الصفات، ويوضح جدول الاقتران بين الظاهرتين كما يلي:

الصفة Y / الصفة X	Y_1	Y_2	…	Y_s	المجموع
X_1	f_{11}	f_{12}	…	f_{1s}	$f_{1.}$
X_2	f_{21}	f_{22}	…	f_{2s}	$f_{2.}$
⋮	⋮	⋮	⋮	⋮	⋮
X_r	f_{r1}	f_{r2}	…	f_{rs}	$f_{r.}$
المجموع	$f_{.1}$	$f_{.2}$	…	$f_{.s}$	$f_{..}$

نحسب المقدار B ومنه نحسب معامل التوافق C بالعلاقة التالية:

$$C = \sqrt{\frac{B-1}{B}}, \qquad (4)$$

حيث:

$$B = \frac{(f_{11})^2}{f_{.1} f_{1.}} + \frac{(f_{12})^2}{f_{.2} f_{1.}} + \ldots + \frac{(f_{rs})^2}{f_{.s} f_{r.}}$$

ونوضح ذلك بالمثال التالي:

مثـال (٤): عند دراسة العلاقة بين الرائحة ولون الزهرة لعينة مكونة من ٣٠ زهرة كانت لدينا النتائج التالية:

الرائحة Y / اللون X	بدون رائحة	له رائحة	المجموع
أصفر	6	4	10
أبيض	7	2	9
أحمر	6	5	11
المجموع	19	11	30

احسب معامل التوافق C بين اللون ورائحة الزهور.

الحل: نحسب قيمة B كالتالي:

$$B = \frac{6^2}{19\times10} + \frac{7^2}{19\times9} + \frac{6^2}{19\times11} + \frac{4^2}{11\times10} + \frac{2^2}{11\times9} + \frac{5^2}{11\times11}$$

$$B = 1.05$$

ويكون معامل التوافق C كالتالي:

$$C = \sqrt{\frac{B-1}{B}}$$

$$C = \sqrt{\frac{1.05-1}{1.05}}$$

$$C = 0.22$$

ويلاحظ أن قيمة معامل التوافق تبين مقدار قوة الارتباط وهى ضعيفة فى هذا المثال.

٥-٤ خط الانحدار البسيط

لقد سبق لنا دراسة العلاقة بين متغيرين (x, y) وإيجاد معامل الارتباط بينهما بعدة طرق وذلك لقياس قوة الارتباط وإتجاه العلاقة بينهما (طردية – عكسية) كما فى معاملى ارتباط بيرسون وسبيرمان ومدى قوة العلاقة كما فى حالة معاملى الاقتران والتوافق. وفيما يلى نبحث عن إيجاد معادلة رياضية تمثل أفضل توفيق لخط مستقيم يعبر عن البيانات فى شكلها الخطي. والغرض من إيجاد معادلة خط الانحدار هو التنبؤ بقيمة المتغير التابع لقيمة محددة من قيم المتغير المستقل. وتسمى العلاقة بين المتغير المستقل x والمتغير التابع y بمعادلة خط الانحدار البسيط. وعليه فإذا كان x متغيرا مستقلاً، y متغيرا تابعا فإن المعادلة التى نحصل عليها تسمى بمعادلة خط انحدار y على x وهى على الصورة التالية:

$$y = a + bx, \qquad (5)$$

حيث يعرف a على أنه ثابت الانحدار (الجزء المقطوع من محور y) ويحسب من خلال العلاقة:

$$a = \bar{y} - b\bar{x}, \qquad (6)$$

كذلك يعرف b بمعامل انحدار y على x ويحسب من خلال العلاقة التالية:

$$b = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}, \qquad (7)$$

مثال (٥): اوجد معادلة خط انحدار درجات الإحصاء y على درجات الرياضيات x فى مثال (١).

الحل: نقوم بتلخيص الحسابات من خلال الجدول التالي:

x	y	xy	x^2	y^2
15	13	195	225	169
7	9	63	49	81
17	19	323	289	361
15	15	225	225	225
10	11	110	100	121
9	8	72	81	64
14	16	224	196	256
10	11	110	100	121
97	102	1322	1265	1398

$$b = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

$$b = \frac{1322 - \frac{97 \times 102}{8}}{1265 - \frac{(97)^2}{8}} = 0.96$$

$$a = \bar{y} - b\bar{x}$$

$$a = \frac{102}{8} - 0.96(\frac{97}{8})$$

$$a = 1.11$$

أى أن معادلة خط انحدار y على x هي:

$$y = 1.11 + 0.96x$$

تمارين

١- فى أحد أماكن بيع السيارات كانت المبيعات كالتالي:

عمر السيارة x	3	2	1	1	5	6	1	4
ثمن البيع y	31	44	60	70	18	17	71	29

- اوجد معامل الارتباط بين عمر السيارة (بالسنوات) وثمن البيع (بـآلاف الجنيهـات) بطريقـة بيرسون.
- اوجد خط انحدار y على x.

٢- الجدول التالى يمثل الدخل x والإنفاق y لمجموعة من الأسر بمئات الجنيهات.

x	56	66	42	44	38	27	39	40
y	31	38	27	22	19	25	20	28

- اوجد معامل ارتباط بيرسون وسبيرمان للدخل والإنفاق.
- اوجد خط انحدار y على x.
- اوجد قيمة الإنفاق عندما يصبح الدخل ٦٠٠٠ جنيه.

٣- البيانات التالية تمثل تقديرات ثمانية من الطلاب فى مادتى الكيمياء والطبيعة.

الكيمياء	A	B	D	E	C	D	E	B
الطبيعة	A	C	E	D	C	D	E	B

اوجد معامل الارتباط المناسب لتقديرات الكيمياء والطبيعة.