
الوحدة الخامسة
الإرتباط والإنحدار الخطي

محتويات الوحدة

الموضوع	الصفحة
المقدمة	128
تمهيد	128
اهداف الوحدة	129
1. الإرتباط الخطي البسيط	130
2. معامل الارتباط الخطي البسيط	137
3. معامل ارتباط الرتب	144
4. معامل الاقتران	147
5. معامل التوافق	149
6. الإنحدار الخطي البسيط	152
الخلاصة	163
لمحة مسبقة عن الوحدة التالية	163

المقدمة

تمهيد

عزيزي الدارس،

مرحباً بك في الوحدة الخامسة من مقرر الإحصاء الإداري، تتألف هذه الوحدة (الإرتباط والإنحدار الخطي) من ستة أقسام رئيسة. شرحنا في القسم الأول المفهوم الإحصائي للارتباط والعلاقة الدالية والإحصائية بين متغيرين (ظاهرتين) وشكل الانتشار بينما تناولنا في القسم الثاني معامل الارتباط الخطي البسيط من بيانات مبوبة وغير مبوبة ، تحدثنا في القسم الثالث عن معامل ارتباط الرتب وفي القسمين الرابع والخامس تناولنا معامل الاقتران ومعامل التوافق على التوالي.

وختاماً تحدثنا في القسم الأخير عن الإنحدار الخطي البسيط من خلال معادلة خط الإنحدار وكيفية رسم خط الإنحدار والعلاقة بين معامل الارتباط ومعامل الانحدار.

عزيزي الدارس ، تحتاج في حياتك العملية إلى معرفة العلاقة بين متغيرين أو أكثر كي تتنبأ بقيمة واتجاه أحدهما إذا علمت قيمة المتغير الآخر ، كأن تعرف التحصيل الدراسي أو العلاقة بين زيادة الدخل الشهري وزيادة ساعات العمل، ولا شك أن دراسة المقاييس الاحصائية السابقة تعطيك تنبأً علمياً دقيقاً في هذا الجانب.

ستجد بين ثنايا هذه الوحدة تعاريف وأسئلة وتمارين تعينك على الفهم الجيد لهذا الموضوع.

أرجو أن تستفيد من دراستك لها كما آمل منك مساهمتك في تقييم هذه الوحدة علمياً بما يدفع في زيادة قيمتها العلمية ولك من الله التوفيق.

أهداف الوحدة



عزيزي الدارس

بعد فراغك من دراسة هذه الوحدة ينبغي أن تكون قادراً على أن:

1. **تكتب** تعريفاً شاملاً وموجزاً لكل من المقاييس الإحصائية الواردة في محتويات هذه الوحدة.
2. **تحل** (مستخدماً أي من هذه المقاييس) مسألتين من ثلاثة مسائل إحصائية حلاً صحيحاً في حدود ستة دقائق.
3. **تفرق** بين استخدامات المقاييس الإحصائية الواردة في هذه الوحدة.
4. **تستخدم** ثلاثة من المقاييس الإحصائية المشار إليها في هذه الوحدة في التنبؤ بعلاقات مستقبلية تخصك لهذا الشهر.
5. **تقارن** العلاقة بين معامل الارتباط ومعامل الإنحدار.

1. الإرتباط الخطي البسيط

Simple Linear Correlation

عزيزي الدارس،

عند دراستنا للمقاييس الإحصائية المختلفة في الوحدة السابقة كان إهتمامنا ينصب على دراسة متغير واحد وهو المتغير تحت الدراسة . ولكن في الحياة العملية قد يصادف الباحث أوضاعاً متعدده . فقد تكون مفردات العينة محل الدراسة عبارة عن أزواج من القيم لظاهريتن مختلفتين أو أكثر ويود الباحث دراسة العلاقة بينهما.

على سبيل المثال العلاقة بين الراتب الشهري ومدة الخدمة والعلاقة بين الطول والوزن لمجموعة من الطلاب والعلاقة بين كمية السماد المستخدم والإنتاج . كل هذه الظواهر لها علاقات مألوفة وهناك الكثير غيرها .

ومن أهم أهداف الحصول على علاقة بين ظاهرتين هي الوصول إلى معادلة تمكن من التنبؤ بقيمة أحد المتغيرين إذا علمنا قيمة المتغير الآخر . وفي هذا الوحدة نقوم بدراسة أساليب قياس مقدار وإتجاه العلاقة بين متغيرين وقوة الإرتباط بينهما ومن ثم تجديد علاقة رياضية تربط المتغيرين وتمكن من التنبؤ .

أ. المفهوم الإحصائي للإرتباط

لكي نبدأ دراسة الإرتباط لابد من التفرقة بين العلاقة الدالية والعلاقة الإحصائية.

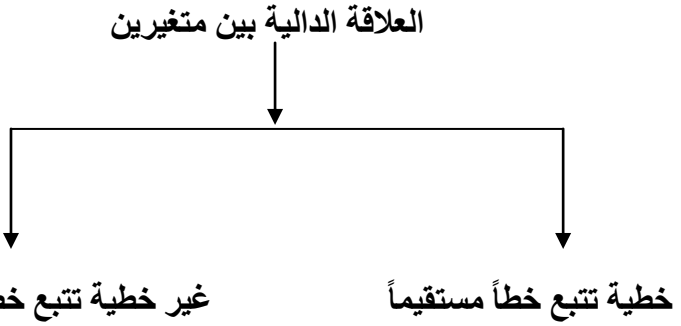
تعريف العلاقة الدالية

العلاقة الدالية بين متغيرين هي علاقة متناظرة تمثلها علاقة رياضية محددة بحيث تحدد قيم أحد المتغيرين تماماً حسب قيم المتغير الآخر.

على سبيل المثال نأخذ العلاقة الدالية بين مساحة الدائرة ونصف قطرها ويعبر عنها رياضياً بالمعادلة :

$$A = \pi r^2$$

حيث تشير A إلى مساحة الدائرة في حين تشير r إلى نصف القطر . فإذا عرف نصف قطر الدائرة أمكن الحصول على مساحتها .

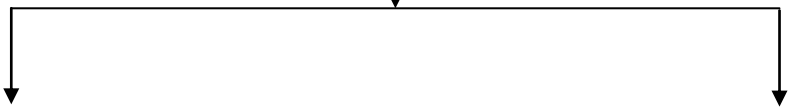


فإذا كانت العلاقة خطية يعبر عنها بمعادلة الخط المستقيم:

$$y = a + bx$$

وتعبر تلك المعادلة عن علاقة خطية دالية بين متغيرين x و y يسمى الأول المتغير المستقل والثاني المتغير التابع . ويمكن إيجاد الخط المستقيم بمعلومية الثوابت a و b . وفي مثل هذه الحالة تقع جميع النقاط (x,y) على الخط المستقيم ويكون الإرتباط بين المتغيرين إرتباطاً خطياً . ويعني ذلك أن أي زيادة أو نقصان في المتغير المستقل (x) تتبعه زيادة أو نقصان في المتغير التابع (y) بمعدل ثابت هو ميل الخط b في معادلة الخط المستقيم .

العلاقة بين متغيرين



طردية ميل الخط المستقيم موجباً (الزيادة في قيمة X تتبعها زيادة في قيمة Y).

عكسية، ميل الخط سالباً (الزيادة في قيمة X يتبعها نقصان في قيمة Y)



1. عرف العلاقة الدالية بين متغيرين؟
2. أكمل الجمل التالية:

أ. العلاقة الدالية بين متغيرين يمكن أن تكون..... وتتبع خطاً مستقيماً كما يمكن أن تكون ..... وتتبع خطاً منحنياً.

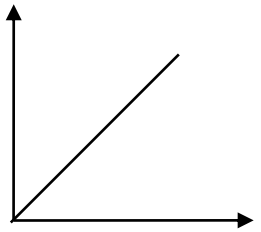
ب. العلاقة بين متغيرين يمكن أن تكون..... حيث يكون ميل الخط موجباً، والزيادة في قيمة X يبتعها ..... في قيمة Y . كما يمكن أن تكون..... حيث يكون ميل الخط سالباً، والزيادة في قيمة X يبتعها ..... في قيمة Y .

ويمكن التعبير عن معادلة الخط المستقيم بالرسم شكل رقم (1) وفيه تخصص محور الاحداثيات الأفقية لقيم x والإحداثيات الرأسية لقيم y المقابلة لها . فإذا كان الخط المستقيم ينطلق من أصل الإحداثيات الرأسية والأفقية يكون ميله يساوي واحد ويصبح الخط المستقيم $y = x$ شكل رقم (1-أ) وتسمى المعادلة معادلة الخط المستقيم المار بنقطة الأصل.

أما إذا كان الخط المستقيم لا يمر بأصل الإحداثيات بل يقطع محور الإحداثيات الرأسية في النقطة (a) عندما تساوي قيمة (x) صفراً فيصبح الخط المستقيم $y = a + bx$ شكل رقم (1–ب)

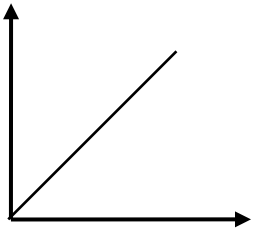
$y = x$

[1–أ]



$y = a + bx$

[1–ب]



شكل رقم (1)

إن العلاقة الخطية الدالية هي علاقة مثالية بحيث أن العلاقة بين متغيرين قد لا تكون بالمعني التام ولكن توجد علاقة من درجة ما بين المتغيرين . وفي هذه الحالة تسمى العلاقة علاقة إحصائية . والعلاقة الإحصائية بين المتغيرين يعبر عنها بالإرتباط بينهما . على سبيل المثال نفترض أننا بصدد إيجاد العلاقة بين متوسطات الدرجات في مادتي الرياضيات X والإحصاء Y في أحد الفصول الدراسية والتي كان توزيعها كما يلي :

جدول (1)

الرياضيات	65	63	67	64	68	62	70	66	68	67	69	71
الإحصاء	68	66	68	65	69	66	68	65	71	67	68	70

وبنظرة سريعة إلى هذه الدرجات يظهر أن هناك علاقة بين درجات الطلاب في هاتين المادتين بإعتبار أن الدرجة إذا كانت مرتفعة في الرياضيات كانت أيضاً مرتفعة في الإحصاء . ولكن هذه العلاقة ليست علاقة سببية بمني أنه إذا كانت درجة الطالب في الرياضيات مرتفعة يصبح من الضروري إرتفاع درجته في الإحصاء ، بل هذه العلاقة إحصائية بحيث أن هناك إتجاه عام لإرتفاع درجات الإحصاء مع إرتفاع درجات الرياضيات .

نشاط



البيانات التالية تمثل حركة القروض والودائع بملايين الدينارات في أحد البنوك التجارية:

القروض (X)	75	89	95	84	72
الودائع (Y)	92	112	116	90	85

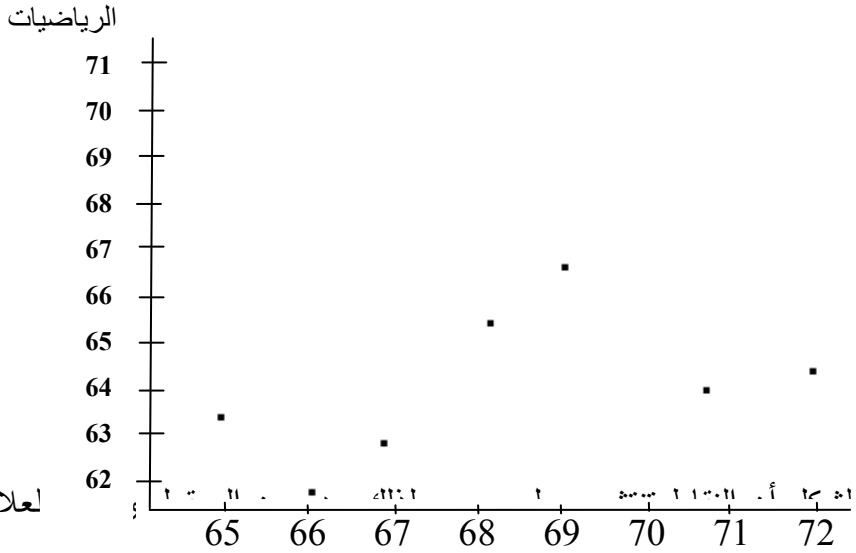
جد معادلتي خطي إنحدار (Y/X) و (X/Y) ثم استخدم النتائج لحساب معامل الإرتباط بين المتغيرين X و Y.

ب – شكل الإنتشار : *Scatter Diagram*

يمكن بيانياً رصد جميع أزواج الدرجات في المثال السابق بحيث يخصص لكل زوج منهما نقطة بالنسبة إلى محورين متعامدين يمثل أحدهما درجات الرياضيات بينما يمثل المحور الثاني درجات الإحصاء فنحصل على ما يسمى بشكل الإنتشار شكل رقم (2) .

شكل رقم (2)

شكل الإنتشار للعلاقة بين درجات الطلاب في الرياضيات والإحصاء



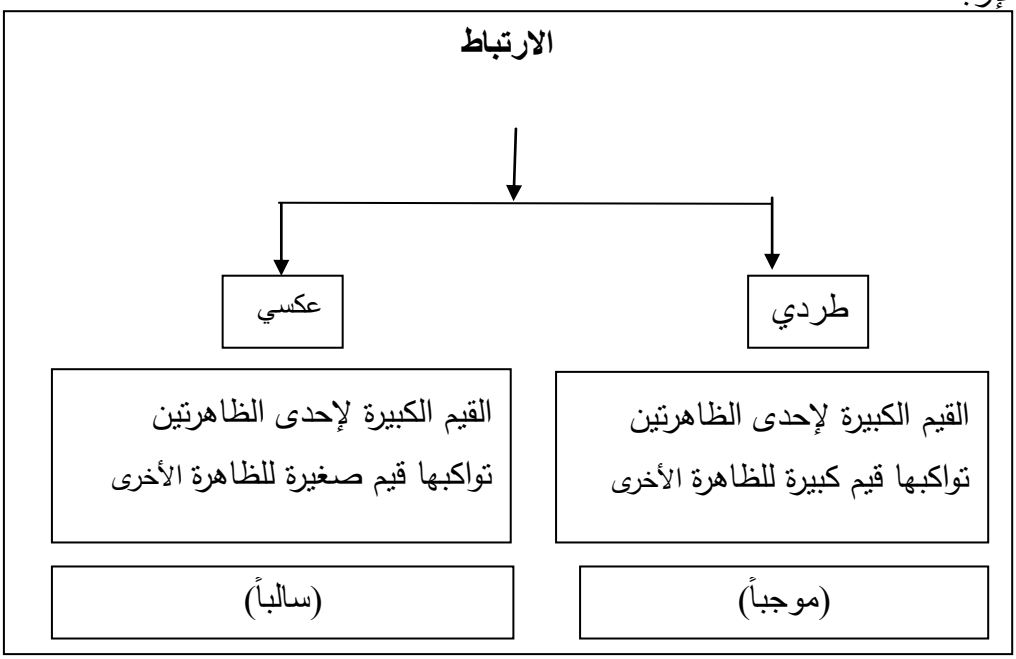
لعلاقة ونلا

بين درجات الرياضيات والإحصاء بالإرتباط ومن ثم التعبير عنها بخط مستقيم يمر ببعض النقاط ويوسط معظمها وتمثله معادله جبرية من النوع :

$$y = a + bx$$

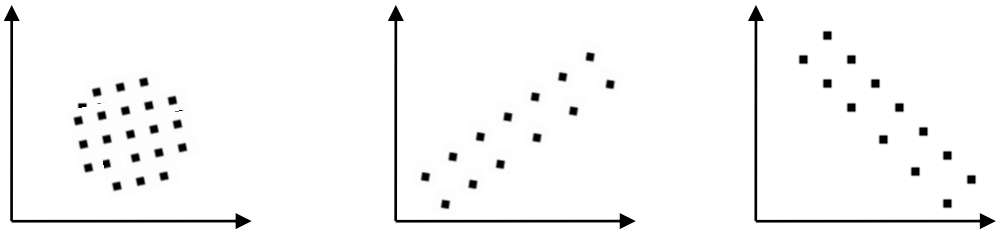
وسنتعرض لهذه المعادلة بالتفصيل في الأقسام القادمة . **يهمنا الآن أن نبين أن شكل الإنتشار يمكن من تكوين فكرة مبدئية عن العلاقة بين متغيرين** . فمن ناحية قد تتجمع النقاط حول إتجاه معين كما رأينا في هذا المثال . بحيث يقال أن المتغيرين بينهما إرتباط من درجة ما . وفي بعض الحالات قد تكون هذه النقاط مبعثرة بلا إتجاه محدد وبالتالي ليس هناك إرتباط بين المتغيرين .

والارتباط بين متغيرين قد يكون طردياً (موجباً) أو عكسياً (سالباً) . الشكل يبين أنواع الإرتباط .



شكل (3)

يبين أنواع الإرتباط



[أ] إرتباط عكسي [ب] إرتباط طردي [ج] لايوجد إرتباط

والإرتباط إما أن يكون قوياً أو ضعيفاً . وإحصائياً يقاس الإرتباط بمقياس رقمي يتراوح بين [1−] و [1+] بحيث تحدد الإشارة الجبرية إتجاه الإرتباط موجباً أو سالباً بينما

تعكس القيمة العددية درجة الإرتباط وقوته . والمقياس الذي يوضح درجة وإتجاه الإرتباط يسمى بمعامل الإرتباط . وسنكتفي هنا بدراسة معامل الإرتباط الخطي . وهناك أكثر من طريقة لحساب معامل الإرتباط الخطي والتي تختلف بإختلاف البيانات المتعلقة بالظاهرتين تحت الدراسة وسنقوم الآن بشرح أهم هذه المعاملات .

أسئلة تقويم ذاتي

1. أرسم معادلة الخط المستقيم في الحالات التالية:
   أ. عندما يمر بنقطة الأصل
   ب. عندما لا يمر بنقطة الأصل
2. أشرح كيف يتم تكوين شكل الانتشار؟
3. بين أنواع الإرتباط وكيف يتم قياسه؟

2. معامل الإرتباط الخطي البسيط

Simple Linear Correlation

ويطلق عليه معامل إرتباط بيرسون *Pearson* ، بالنسبة للمتغيرين y ، x يحسب إرتباط بيرسون على أنه متوسط مجموع حواصل ضرب القيم المعيارية للمتغيرين ويرمز له بالرمز r ، ويعبر عنه رياضياً كما يلي :

$$r = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{n\sigma_x \sigma_y}$$

وهذا المقياس يعبر عن كل خصائص الإرتباط الخطي البسيط من ناحية القوة والإتجاه فهوعبارة عن مجموع ضرب الإنحرافات للمتغيرين x ، y عن وسطهما الحسابي مقسوماً على إنحرافهما المعياري . فعندما يكون إنحراف x سالباً وإنحراف y سالب يكون الإرتباط موجب أيضاً . أما إذا كان أحد الإنحرافات موجب والثاني سالب فيكون الإرتباط سالب .

للأغراض الحسابية يمكن تحوير المعادلة السابقة إلى عدة أشكال :

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\{N\sum X^2 - (\sum X)^2\}\{N\sum Y^2 - (\sum Y)^2\}}}$$

والمعادلة السابقة هي أكثر المعادلات إستخداماً ذلك أنها تستخدم القيم الأصلية للمتغيرين x , y دون اللجوء إلى حساب الإنحرافات والوسط الحسابي والإنحراف المعياري للمتغيرات .

نوضح في الجدول التالي خطوات حساب معامل الإرتباط البسيط بين درجات الرياضيات والإحصاء الوارده في الجدول رقم (1) .

جدول رقم (2)

خطوات حساب معامل الإرتباط البسيط بين درجات الرياضيات والإحصاء

Y^2	X^2	**XY**	**Y**	**X**	**الإحصاء Y**	**الرياضيات X**
64	25	40	8	5	68	65
36	9	18	6	3	66	63
64	49	56	8	7	68	67
25	16	20	5	4	65	64
81	64	72	9	8	69	68
36	4	12	6	2	66	62

64	100	80	8	10	68	70
25	36	30	5	9	65	66
121	64	88	11	8	71	68
49	49	49	7	7	67	67
64	81	72	8	9	68	69
100	121	110	10	11	70	71
$\sum Y^2 = 729$	$\sum X^2 = 618$	$\sum XY = 647$	$\sum Y = 91$	$\sum X = 80$		

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\{N\sum X^2 - (\sum X)^2\}\{N\sum Y^2 - (\sum Y)^2\}}}$$

بالتعويض في المعادلة أعلاه نجد أن معامل الإرتباط الخطي البسيط :

$$r = \frac{(12)(657) - (80)(91)}{\sqrt{\{(12)(618) - (80)^2\}\{(12)(729) - (91)^2\}}}$$

نلاحظ أننا قمنا بطرح مقدار ثابت من كل درجة من الإحصاء والرياضيات لتسهيل العمليات الحسابية ، وذلك لا يؤثر على معامل الإرتباط حيث أنه لا يعتمد على إختيار نقطة الأصل .

ونلاحظ أن قيمة r أعلى من 0.5 مما يدل على أن هنالك علاقة طردية موجبة ذات إرتباط متوسط . وتجدر الإشارة إلى أنه لاتوجد حدود فاصلة تحدد قوة وضعف الإرتباط ولكن من الخبرة يمكن تحديد ذلك .

ومن أهم خصائص معامل الإرتباط الخطي لبيرسون أنه لا يتأثر بإضافة أو طرح أي مقدار ثابت بالنسبة لجميع قيم الظاهرتين . كما يتمتع معامل الإرتباط بهذه الخاصية في حالة الضرب والقسمة . ويمكن استخدام هذه الخواص في حساب معامل الإرتباط كما بينا

في المثال السابق . ومن الإحصائيات المتوفره هناك حدود تقريبية لقيم r يتم العمل بها ، انظر جدول رقم (3) ، وتتغير الإشارة عندما تكون r سالبة .

جدول رقم (3)

درجة قوة الإرتباط *

قيمة r	درجة قوة الإرتباط
<0.3	لا يوجد إرتباط يذكر
0.3 – 0.49	إرتباط ضعيف
0.5 – 0.69	إرتباط متوسط
0.7 – 0.89	إرتباط قوي
0.9 – 1.00	إرتباط قوي جداً

*ويمكن أن تكون هذه القيم سالبة في حالة الإرتباط العكسي .

معامل الإرتباط الخطي لبيرسون من البيانات المبوبة :

في هذه الحالة ندخل التكرارت f على معادلة معامل الإرتباط الخطي البسيط لتأخذ الشكل :

$$r = \frac{N\sum f_{xy}u_x u_y - (\sum f_x u_x)(\sum f_y u_y)}{\sqrt{\{N\sum f_x u_x^2 - (\sum f_x u_x)^2\}\{n\sum f_y u_y^2 - (\sum f_y u_y)^2\}}}$$

حيث :

u_x = انحرافات مراكز فئات X عن وسط فرضي .

u_y = انحرافات مراكز فئات Y عن وسط فرضي .

N = مجموع التكرارات .

f_x = التكرارات (الهامشية) المقابلة للفئة x .

f_y = التكرارات (الهامشية) المقابلة للفئة y .

$f_x u_x$ = تكرار الفئة x مضروباً في الإنحراف المقابل للمتغير x .

$f_y u_y$ = تكرار الفئة y مضروباً في الإنحراف المقابل للمتغير y .

$f_{xy} u_x u_y$ = حاصل ضرب إنحرافي x و y في التكرار المشترك لفئتيهما.

مثال (1)

أوجد معامل الإرتباط الخطي بين درجات الرياضيات y ودرجات الإحصاء x لمجموعة مكونة من 30 طالب في الجدول التالي :

جدول رقم (4)

حساب معامل الارتباط الخطي

y / x	−50	−60	−70	−80	99−90	**المجموع**
−50	3					3
−60	1	4				5
−70	2	8				10
−80			1	6	1	8
99−90					4	4
المجموع	6	12	1	6	5	**100**

وفي الجدول يمثل العمود الأول فئات المتغير x والسطر الأول فئات المتغير y . أما الأرقام في الخلايا داخل الجدول فتمثل التكرارات المشتركة للفئتين المقابلتين . فمثلاً الرقم (3) في الخلية التي تقع في تقاطع الصف الثاني والعمود الثاني يمثل عدد الوحدات التي تتراوح قيمة x فيها بين 50 إلي أقل من 60 وقيمة y بين 50 إلي أقل من 60 ... وهكذا

ولحساب معامل الإرتباط الخطي لبيرسون نقوم بتنظيم خطوات الحل في جدول (5–6) على النحو التالي :

جدول رقم (5)

حساب معامل الإرتباط الخطي من بيانات مبوبة

$f_{xy}u_xu_y$	$f_xu_x^2$	f_xu_x	u_x	f_x	−94.5	−84.5	−74.5	64.−5	−54.5	Y X
12	12	−6	−2	3					12 3	−54.5
6	5	−5	−1	5				4 4	2 1	−64.5
0	0	0	0	10				0 8	0 2	−74.5
8	8	8	1	8	2 1	6 6	0 1			−84.5
16	16	8	2	4	16 4					−94.5
42	41	5		30	5	6	1	12	6	f_y
					2	1	0	−1	−2	u_y
				−8	10	6	0	−12	−12	f_yu_y
				62	20	6	0	12	24	$f_yu_y^2$
				42	18	6	0	4	14	$f_{xy}u_xu_y$

ونلاحظ في الجدول أن القيم المحصورة بين المربعات داخل الخلايا تمثل التكرار المشترك للفئتين المقابلتين مضروباً في حاصل ضرب إنحرافيهما عن الوسط الفرضي. فمثلاً العدد 12 في العمود الأول تم حسابه بضرب التكرار المشترك للفئتين الأولتين X و Y وهو (3) في إنحرافيهما وهما 2− و 2− .

$$f_{xy}u_xu_y = -2(-2)(3) = 12$$

ويحتوي الجدول على الأعمدة التالية :

u_x = الإنحرافات بطريقة الترميز لمراكز الفئات بالنسبة للمتغير x ، حيث نضع (0) مقابل أي فئة (يفضل الوسطي) ثم 1− ، 2− ، ... للفئات قبلها بالترتيب ، و 1+ ، 2+ ، ... للفئات بعدها . فمثلاً في حالة X وضعنا الصفر مقابل الفئة 74.5− 84.5 مما نتج عنه أن الإنحراف المقابل للفئة 54.5− فمثلاً ، 2− .

u_y= الإنحرافات بطريقة الترميز لمراكز الفئات بالنسبة للمتغير y .

باقي الأعمدة تم حسابها مباشرة عن طريق ضرب التكرارات المشتركة لكل فئتين متقابلتين في إنحرافاتها. وقد تم حساب $\sum f_x u_x u_y$ أو $\sum f u_x u_y$ بواسطة جمع الأرقام الموجودة في المربعات الجانبية بكل خلية بالنسبة لكل صف ولكل عمود . ونلاحظ أن المجاميع الكلية في الصف الأخير والعمود الأخير متساوية وهذا أمر مهم بالنسبة لمراجعة حساب الإرتباط .

ومن الجدول نحصل على :

$$r = \frac{N\sum f_{xy}u_x u_y - (\sum f_x u_x)(\sum f_y u_y)}{\sqrt{\{(N\sum f_x u_x^2 - (\sum f_x u_x)^2\}\{n\sum f_y u_y^2 - (\sum f_y u_y)^2)\}}}$$

$$r = \frac{(30)(42) - (-8)(5)}{\sqrt{\{(30)(62) - (-8)^2\}\{(30)(41) - (5)^2\}}}$$

$$r = \frac{1300}{1471.1} = 0.884$$

أي أن الإرتباط طردي قوي بين درجات الرياضيات و الإحصاء .

نشاط



الجدول التالي يمثل تكلفة السماد وعائد الإنتاج بالآف الدينارات السودانية:

تكلفة السماد (X)	9	10	9	10	4	5	7
المبيعات (Y)	160	180	150	140	120	120	140

أ –أرسم شكل الإنتشار للمتغيرين X و Y.

ب أحسب معامل الإرتباط الخطي لبيرسون .

3. معامل إرتباط الرتب *Rank Correlation*

في الحالات التي يكون فيها أحد المتغيرين أو كليهما ترتيبياً ، فإننا لا نستطيع إستخدام معامل الإرتباط الخطي البسيط . لذلك نشأت الحاجة إلى **إيجاد مقياس يستخدم في مثل هذه الحالة ، ومن أمثلة البيانات الترتيبية ترتيب عدد من العاملين حسب درجة رضائهم الوظيفي أو حماسهم للعمل ويعتمد معامل إرتباط الرتب لسبيرمان على فروق الرتب المتناظرة بين قيم المتغيرين ، والذي يحسب من المعادلة :**

$$r = 1 - \left[\frac{6\sum d^2}{n(n^2 - 1)} \right]$$

حيث:

d = فروق الرتب بين قيم المتغيرين ، n = عدد أزواج القيم.

والبيانات قد تكون في شكل رتب أصلاً ، أو قد تكون قيماً تحتاج لترتيب . في الحالة الأخيرة تعطى رتب لقيم كل من المتغيرين على حدة حيث نأخذ أًصغر القيم للمتغير الرتبة (1) والتي تليها في الكبر الرتبة (2) وهكذا . وكذلك ترتب قيم المتغير الآخر . وإذا تساوت قيمتان أو أكثر عند ترتيب قيم متغير نعطي كل منهما متوسط رتبها إذا لم تكن متساوية .

فمثلاً إذا تساوت قيم كل المفردات التي كان سيكون ترتيبها لو أنها مختلفة 6، 5 ، 4 فإن كل منها تعطى الرتبة 5 . وبالمثل إذا تساوت قيمتا المفردتين اللتان ترتيبهما 8 ، 7 فتعطى كل منهما الرتبة 7.5 وهكذا . ثم نكوّن جدولاً نضع فيه قيم المتغيرين ورتب كل منهما ثم نحسب الفروق بين الرتب ثم مربعاتها ونعوض في المعادلة الأخيرة لنحصل على معامل الإرتباط .

مثال (2)

حصل 10 طلاب على التقديرات الآتية في مادتي الإحصاء والرياضيات .

جدول رقم (6)

رقم الطالب	الإحصاء	الرياضيات
1	C	B
2	F	C
3	C+	C
4	A	B
5	C	C
6	D	C+
7	B	C
8	C+	A
9	D	F
10	C	D

أحسب معامل الإرتباط .

ونبين خطوات الحل في الجدول التالي .

جدول رقم (7)

حساب معامل إرتباط الرتب

رقم الطالب	الإحصاءY	الرياضياتX	رتب X	رتب Y	فروق الرتب d	d^2
1	C	B	6	2.5	3.5	12.25
2	F	C	10	6.5	3.5	12.25
3	C+	C	3.5	6.5	3.0	9
4	A	B	1	2.5	1.5	2.25
5	C	C	6	6.5	0.5	0.25
6	D	C+	8.5	4	4.5	20.25
7	B	C	2	6.5	4.5	20.25
8	C+	A	3.5	1	2.5	6.25
9	D	F	8.5	10	1.5	2.25
10	C	D	6	9	3.0	9
المجموع						94

من هذا الجدول نجد أن معامل الإرتباط :

$$r = 1 - \left[\frac{6\sum d^2}{n(n^2-1)} \right]$$

$$r = 1 - \left[\frac{6(94)}{10(100-1)} \right]$$

$$r = 1 - \frac{564}{999} = 0.43$$

ويلاحظ أنه وفي هذا المثال الإرتباط طردي ضعيف بين درجات الطلاب في الإحصاء والرياضيات .

ومن أهم ما يمتاز به هذا القياس للإرتباط أنه لا يصلح فقط للبيانات الترتيبية وإنما يمكن إستخدامه أيضاً لقياس الإرتباط في حالة البيانات الرقمية . إلا أنه يؤخذ عليه أنه أقل دقة بالمقارنة بمعامل إرتباط بيرسون في هذه الحالة.

أسئلة تقويم ذاتي



1. عرف معامل ارتباط بيرسون وأكتب معادلته؟
2. ما هي أهم خصائص معامل ارتباط بيرسون؟
3. على ماذا يعتمد معامل ارتباط الرتب لسبيرمان وكيف يتم حسابه؟

نشاط



أحسب معامل إرتباط سيبرمان للرتب من واقع البيانات الواردة في الجدول التالي والذي يمثل تقديرات مجموعة من الطلاب في مادتين:

تقديرات المادة الأولى(X)	D	A	D	F	B	F	D
تقديرات المادة الثانية(Y)	D	C	B	D	A	D	C

4. معامل الإقتران Contingency Coefficient

في بعض الأحيان قد تصادفنا بيانات وصفية ليس في طبيعتها صفة للترتيب أي لا يمكن وضع رتب لها أو لا معنى للرتب . فإذا أخذنا مثلاً فصائل الدم نجد الفصائل A ، B وغيرها أو المدخنين فنجد أن التقسيم يمكن أن يكون يدخن أم لا . وهناك الكثير من الأمثلة التي يكون للمتغيرين فيها صفتان أولي وثانية .

ويمكن التعبير عن جدول الإقتران بالجدول التالي .

جدول (6)

إقتران بين الصفات

X_2	X_1	**X** / **Y**
B	A	Y_1
D	C	Y_2

ونلاحظ أن A هي عبارة عن التكرارت المشتركة في الصفة الأولى X والصفة الثانية Y وهكذا بالنسبة لبقية خلايا الجدول . ويستخدم الإحصائيون في قياس معامل الإقتران ما يعرف بمعامل ييل (Yule) للإقتران والذي يأخذ الشكل التالي :

$$R_C = \frac{AD - BC}{BC + AD}$$

مثال (3)

أوجد معامل إقتران بين لون العيون ولون الشعر لمجموعة من الأفراد من البيانات في جدول (9-6)

جدول (7)

الإقتران بين لون الشعر ولون العيون

لون العيون x / لون الشعر y	أزرق	عسلي
أشقر	8	6
أسود	7	5

من هذا الجدول يمكن حساب معامل الإقتران مباشرة بتطبيق القانون السابق :

$$R_C = \frac{AD - BC}{BC + AD} = \frac{40 - 42}{42 + 40} = \frac{-2}{82} = -0.024$$

ونلاحظ أنه لا يوجد إرتباط بين لون الشعر ولون العيون إذ أن معامل الإرتباط قريب جداً من الصفر .

نشاط



في تجربة لدراسة أثر التدخين في الإصابة بمرض السرطان كانت النتائج كما يلي:

التكررات المشتركة للتدخين والإصابة بالسرطان

الإصابة / التدخين	أصيب بالمرض	لم يصب بالمرض
يدخن	10	90
لايدخن	4	94

جد مدي علاقة التدخين بالإصابة بالسرطان .

5. معامل التوافق

في حالة تعدد صفات المتغيرات لا يمكن تطبيق قانون ييل . على سبيل المثال إذا كانت البيانات الخاصة بالعيون تحتوي على أكثر من صفتين (أسود ، أزرق ، عسلي ،.....) أو لون الشعر يحتوي أكثر من صفتين أيضاً (أشقر ، بني ، أسود) في مثل هذه الحاله لابد وأن يشتمل جدول الإقتران على جميع الصفات الوارده بالنسبة للمتغيرين في جدول مبسط كما يلي :

جدول إقتران بين عدد من الصفات

المتغير x / المتغير y	X_1	X_2	$X3$	**المجموع**
Y_1	C_{11}	C_{12}	C_{13}	C_1
$Y2$	C_{21}	C_{22}	C_{23}	C_2
$Y3$	C_{31}	C_{32}	C_{33}	C_3
المجموع	$\sum C_1$	$\sum C_2$	$\sum C_3$	$\sum C$

جدول رقم (8)

حيث ……..j=1,2,3, ……..i=1,2,3,

وفي هذه الحالة يعرف معامل الإقتران بمعامل التوافق ويرمز له بالرمز R ويمكن حساب قيمته باستخدام العلاقة التالية :

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

حيث أن $C = \sum C_1$

تدريب(1)



أوجد معامل التوافق بين الطول والوزن لعينة مكونة من 45 طالباً بإحدي الجامعات باستخدام البيانات التالية .

الطول X / الوزن Y	ضخم	متوسط	نحيف	المجموع
طويل	7	5	3	15
متوسط	6	3	6	15
قصير	6	5	4	15
المجموع	17	13	15	45

لإيجاد معامل الإقتران نحسب أولاً قيمة C كالآتي :

$$\sum C_1 = \frac{(7)^2}{15\times17} + \frac{(5)^2}{15\times13} + \frac{(3)^2}{15\times15} = 0.360$$

$$\sum C_2 = \frac{(6)^2}{15\times17} + \frac{(3)^2}{15\times13} + \frac{(6)^2}{15\times15} = 0.347$$

$$\sum C_3 = \frac{(4)^2}{15\times17} + \frac{(5)^2}{15\times13} + \frac{(6)^2}{15\times15} = 0.347$$

$$C = \sum C_1 + \sum C_2 + \sum C_3 = 1.058$$

$$R_C = \sqrt{\frac{C-1}{C}} = \sqrt{\frac{1.058-1}{1.058}} = 0.23$$

لا يوجد إرتباط يذكر بين أطوال وأوزان الطلاب .

6. الإنحدار الخطي البسيط

Simple Linear Regression

كما ذكرنا فيما سبق أنه من أهم أهداف البحث العلمي الحصول على علاقة بين المتغيرات تمكن من التنبؤ في المستقبل . وذلك باستنباط قيم أحد المتغيرين بدلالة المتغير الآخر . وأوضحنا من توزيع النقاط بشكل الإنتشار أن العلاقة الإحصائية بين ظاهرتين . ربما تأخذ شكل الخط المستقيم ويطلق عليها الإنحدار الخطي البسيط .

وتشير معادلة خط الإنحدار إلى أنحدار أحد المتغيرين وليكن Y ويسمي متغير تابع على المتغير الآخر وليكن X ويسمى متغير مستقل ، وتسمي العلاقة عندئذ خط إنحدار Y على X ويرمز له بـ E(Y/X) وكذلك معادلة خط إنحدارالمتغير X على المتغير Y ويرمز له بـ E(X/Y).

- **معادلة خط إنحدار Y على X :**

تكتب هذه المعادلة على الصورة الآتية :

$$E(Y/x) = \alpha + \beta x$$

حيث α, β ثابتان ، ويتحدد الإنحدار متى أمكن تحديدهما و في نموذج الانحدار البسيط لـ Y على X نفترض أن أي قيمة لـ Y ، مثلاً y ترتبط بقيمة معينة لـ X ، مثلاً x ، بالمعادلة :

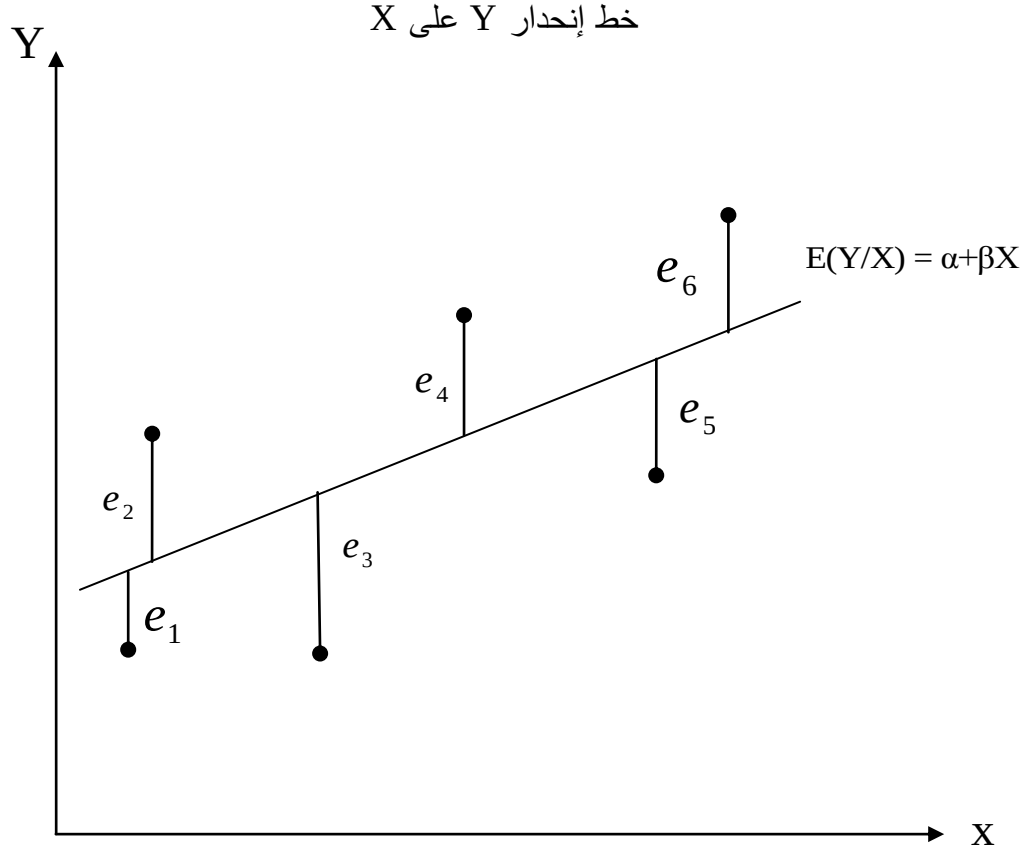
$$y = \alpha + \beta x + e$$

حيث e متغير عشوائي .

ويمكن توضيح هذه الصورة بالشكل التالي :

شكل رقم (5)

خط إنحدار Y على X



في الشكل النقاط تمثل 6 أزواج مرتبة (x_6,x_n) ….., (x_2,x_2) , (x_1,y_1) من المشاهدات . أما الخط المستقيم فيمثل خط انحدار Y على X . أما الخطوط الرأسية فتمثل الأخطاء العشوائية $e_1,e_2,\ldots..,e_6$ أي انحرافات القيم الفعلية لـ y عن القيم المقابلة بالخط .

نلاحظ أن الخط المستقيم يمر ببعض النقاط ويتوسط معظمها والمقصود منه تمثيل العلاقة الخطية بين المتغيرين أصدق تمثيل. **والسؤال هو كيف يتم رسم مثل هذا الخط ؟**

أ- التمهيد باليد :

وهنا يتم تمهيد الخط المستقيم المناسب لنقاط شكل الإنتشار للبيانات الخاصة بالمتغيرين X ، Y . ويعاب على هذه الطريقة أنها تقريبية جداً وتخضع للتقدير الشخصي . ولهذا كان لابد من إيجاد طريقة لتوفيق خط الإنحدار بحيث لا تعتمد على إنطباع الأشخاص وإنما على البيانات الخاصة بالمتغيرين X و Y .

ب – طريقة المجموعات :

وفي هذه الطريقه يتم قسمة البيانات الخاصة بالمتغيرين X ، Y إلى مجموعتين ثم يؤخذ المتوسط لمفردات النتغير Y في المجموعة الأولى ويسمى $\overline{Y}_1$ والمتوسط لمفردات Y في المجموعة الثانية ويسمى $\overline{Y}_2$ وتطبق نفس هذه الطريقة على المتغير X لنحصل على $\overline{X}_1, \overline{X}_2$ وبهذه الطريقة يكون لدينا مجموعتان إحداثياتهما $(\overline{X}_1,\overline{Y}_1),(\overline{X}_2,\overline{Y}_2)$ ، وعليه يمكن حساب الثابت β الذي يمثل ميل الخط المستقيم من العلاقة :

$$\beta = \frac{\overline{Y}_2 - \overline{Y}_1}{\overline{X}_2 - \overline{X}_1}$$

وبالتعويض في إحدي المعادلتين يمكن إيجاد الثابت α.

مثال(5)

جد خطاً مستقيماً يمثل البيانات التالية .

X	2	3	5	7	9	10
Y	1	3	7	11	15	17

الحل

أولاً: نقسم قيم X إلى مجموعتين هما (2،3،5) و (7،9،10).

ثم نقسم قيم Y إلى مجموعتين أيضاً هما (1،3،7) و (17، 11،15) .

ثانياً : نحسب الوسط الحسابي لكل مجموعة من المجموعات الأربعة كما يلي :

$$\overline{X}_1 = \frac{2+3+5}{3} = 3.33$$

$$\overline{X}_2 = \frac{7+9+10}{3} = 8.67$$
$$\overline{Y}_1 = \frac{1+3+7}{3} = 3.67$$
$$\overline{Y}_2 = \frac{17+15+11}{3} = 14.33$$

ثالثاً : الإحدثيات هي : $(\overline{X}_1, \overline{Y}_1) = (3.33,3.67), (\overline{X}_2, \overline{Y}_2) = (8.67,14.33)$

وعليه فإن ميل الخط الخط β يساوي :

$$\beta = \frac{\overline{Y}_2 - \overline{Y}_1}{\overline{X}_2 - \overline{X}_1} = \frac{14.33 - 3.67}{8.67 - 3.33} = 2.00$$

وبتعويض ميل المعادلة مع أي نقطتين X ، Y نجد أن $\alpha = -3$ والمعادلة المطلوبة هي

$$\boxed{Y = 2X-3}$$

وهذه الطريقة وإن كانت تعتمد على بيانات المتغيرات الأصلية إلا أنها تعتمد على الأوساط الحسابية للنقاط على طرفي التوزيع وهي تعطي نتائج غيرملائمة في حالة التوزيعات المتطرفة أو إذا كانت المفردات لاتقبل القسمة على 2 . لتلافي مثل هذه الامور فإننا نلجأ إلى صيغة رياضية لخط الإنحدار تسمى بطريقة المربعات الصغرى .

ج – طريقة المربعات الصغرى :

إذا نظرنا لشكل الإنتشار في شكل رقم (5−6) نجد أن المسافة بين أي نقطة والخط المستقيم هي $e_i = D_i$ حيث (i= 1,2,…n) وتسمي هذه القيم بالإنحرافات ونسبة لأن هذه الإنحرافات تقع على الخط مباشرة أو على جانبيه فهي إما أن تكون سالبة أو موجبة أو قيمتها صفر، وعليه يمكن الإنطلاق من الإفتراض أن مجموع هذه الإنحرافات صغير للغاية .

وأساس طريقة المربعات الصغرى هو اعتبار الخط الثاني الذي يطابق النقاط أحسن مطابقة وهو الخط الذي يكون فيه مجموع مربعات انحرافات النقط عنه أصغر ما يمكن .

فإذا كان الخط أو الإنحراف في تمثيل النقطة (X_i, Y_i) عن خط الإنحدار هو e_i فإن :

$$e_i = Y_i - \alpha - \beta X_i$$

وعليه فإن مجموع مربعات الإنحرافات هو :

$$\sum e_i^2 = \sum (Y_i - \alpha - \beta X_i)^2$$

ولكي يكون($\sum e_i^2$) نهاية صغرى بمعنى أن الأبعاد الرأسية للنقطة عن الخط المقترح أصغر ما يمكن فإننا نفاضل ($\sum e_i^2$) بالنسبة لكل من α وβ ونجعل نتيجة كل منهما تساوي الصفر فنحصل على المعادلتين التاليتين :

$$\sum Y_i = n\alpha - \beta \sum X_i$$

$$\sum X_i Y_i = \alpha \sum X_i + \beta \sum X_i^2$$

حيث n عدد أزواج القيم المعلومة .

وتسمى هذه المعادلات بالمعادلات المعتادة ، ويمكن حل المعادلات المعتادة جبرياً لنحصل على مقدرين لقيمتي الثابتين α وβ كما يلي :

$$\hat{\alpha} = \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{n\sum X^2 - (\sum X)^2} = \frac{\sum Y}{n} - \hat{\beta}\frac{\sum X}{n}$$

$$\hat{\beta} = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2}$$

ويسمي الثابت $\hat{\beta}$ عادة معامل إنحدار Y على X والثابت $\hat{\alpha}$ هو الجزء الذي يقطعه خط إنحدار Y على X من المحور Y.

مثال(6)

أوجد معادلة إنحدار Y الإنفاق على X الدخل للبيانات التالية ، ومن ثم جد تقدير الإنفاق عندما يكون الدخل 2500 دينار سوداني .

الحل

نقوم بتفصيل الحل في الجدول التالي :

جدول رقم (9)

معادلة الإنحدار (Y) الإنفاق على (X) الدخل

Y^2	X^2	XY	الإنفاق Y	الدخل X
64	64	64	8	8
81	100	90	9	10
144	144	144	12	12
100	144	120	10	12
100	169	130	10	13
169	225	195	13	15
361	400	380	19	20
$\sum Y^2 = 1019$	$\sum X^2 = 1246$	$\sum XY = 1123$	$\sum Y = 81$	$\sum X = 90$

أولاً نحسب معامل الإنحدار :

$$\hat{\beta} = \frac{(7)(1123) - (90)(81)}{(7)(1246) - (90)^2} = 0.92$$

$$\hat{\alpha} = \frac{(81)(1246) - (90)(1123)}{(7)(1246) - (90)^2} = 0.23$$

وعليه فإن معادلة خط إنحدار (X/Y) تصبح :

$$Y= -0.23 + 0.92\ X$$

وتكون قيمة الإنفاق Y عندما يكون الدخل 2500 دينار هو :

$$Y= -0.23 + 0.92\ (2500) = 2299.77$$

أي أن الإنفاق يساوي : 2300 دينار تقريباً .

- **معادلة خط إنحدار (X/Y)**

بصورة مماثلة لخط إنحدار (Y/X) يمكن توفيق خط إنحدار (X/Y) على الصورة التالية :

$$\boxed{X_i = \pi + \rho Y_i + e_i}$$

حيث Y ، X متغيران ، يمثل X المتغير التابع ويمثل Y المتغير المستقل ، أما π, ρ فهما ثابتا المعادلة تماماً مثل α وβ في معادلة انحدار (Y/X) ، أما e_i فهو المتغير العشوائي .

نلاحظ في هذه المعادلة أن التغير في X ينشأ من مصدرين هما التغير في المتغير المستقل Y والذي يمكن قياسه ثم المتغير العشوائي e والذي لا يمكن قياسه . وكما فعلنا في حالة إنحدار (Y/X) يمكن إستخدام طريقة المربعات الصغري لنحصل على المعادلات المعتادة التاية :

$$\boxed{\sum X_i = n\pi + \rho \sum Y_i}$$

$$\sum X_i Y_i = \rho \sum Y_i + \rho \sum Y_i^2$$

حيث n عدد أزواج القيم المعلومة .

وتسمى هذه المعادلات بالمعادلات المعتادة ، ويمكن حل المعادلات المعتادة جبرياً لنحصل على تقدير لقيمتي الثابتين π, ρ كما يلي :

$$\hat{\pi} = \frac{(\sum X)(\sum Y^2) - (\sum Y)(\sum XY)}{n\sum Y^2 - (\sum Y)^2}$$

$$\hat{\rho} = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum Y^2 - (\sum Y)^2}$$

مثال(7)

جد معادلة خط إنحدار (X/Y) ، من بيانات الدخل والإنفاق في المثال السابق ثم جد تقدير الدخل إذا كان الإنفاق 2000 دينار ؟ .

الحل

لحل هذه المسأله لا نحتاج لجدول آخر حيث أن جميع البيانات المطلوبة لحساب $\hat{\pi},\hat{\rho}$ موجودة في جدول رقم (9–6) ، وعليه فإن :

$$\hat{\pi} = \frac{90(1019) - 90(1123)}{7(1019) - (81)^2} = -16.36$$

$$\hat{\rho} = \frac{7(1123) - 90(81)}{7(1019) - (81)^2} = 1.00$$

وعليه فإن معادلة خط إنحدار (X/Y) هي :

$$X = -16.36 + Y$$

فإذا كان الإنفاق 2000 دينار فإن الدخل يساوي :

$$X = 16.36 + (2000) = 1984$$

ملاحظة :

قد يلاحظ الدارس أن المتغير المستقل في هذا المثال لا يصلح أن يكون كمتغير تابع ، ولكن هنالك اكثير من العلاقات الخطية يمكن تبديل المتغيرات داخلها . وقد أوردنا نفس المثال في معادلة (X/Y) أشرح بعض المفاهيم الإحصائية المتعلقة بين الإرتباط والإنحدار .

ج – العلاقة بين معامل الإرتباط ومعامل الإنحدار :

عندما يكون الإنحدار بين ظاهرتين مستقيماً ، فإن الإرتباط بينهما أيضاً يكون كذلك ، وتكون هناك علاقة رياضية مباشرة بين العاملين نبينها فيما يلي :

أولاً : معادلة إنحدار (Y/X) :

$$\beta = \frac{\sum XY - (\sum X)(\sum Y)}{n\sigma_x}$$

فإذا قمنا بضرب طرفي المعادلة في $\frac{\sigma_x}{\sigma_y}$ نجد أن :

$$\beta\frac{\sigma_x}{\sigma_y}=\frac{\sum XY-(\sum X)(\sum Y)}{n\sigma_y}\Rightarrow r=\beta\frac{\sigma_x}{\sigma_y}\rightarrow *$$

وعليه فإن معامل الإرتباط يساوي معامل الإنحدار مضروباً في الإنحراف المعياري للمتغير X مقسوماً على الإنحراف المعياري للمتغير Y .

ثانياً : معادلة إنحدار (X/Y)

بالمثل إذا قمنا بضرب معامل إنحدار (X/Y) في $\frac{\sigma_x}{\sigma_y}$ نحصل على :

$$r=\rho\frac{\sigma_x}{\sigma_y}\rightarrow **$$

وإذا ضربنا المعادلتين [*] و [**] ببعضهما البعض نحصل على : $r^2=\beta\rho$

أي أن معامل الإرتباط يساوي الجذر التربيعي لمعاملي الإنحدار $r=\sqrt{\beta\rho}$

ويستخدم هذه العلاقة لحساب معامل الإرتباط بمعلومية معاملي خطي الإنحدار .

مثال(8)

جد معامل الإرتباط من بيانات الدخل و الإنفاق في المثال السابق .

الحل

لقد أوجدنا معاملي خطي الإنحدار (Y/X) و (X/Y) حيث كان $\beta=0.92,\rho=1.00$

وعليه فإن معامل الإرتباط r هو :

$$r=\sqrt{\beta\rho}=\sqrt{0.92(1.00)}=\sqrt{0.92}=0.96$$

وهي نفس النتيجة التي توصلنا إليها عند حساب معامل الإرتباط الخطي لبيرسون من البيانات مباشرةً .

نشاط



[5] – في دراسة عن الخصوبة لعينة من الأسر بمحافظة الجزيرة كانت العلاقة بين عمر الأم وعدد الأطفال كما يلي :

عدد الإطفال / العمر	2-	4-	6-	8-10
20-	12			
25-	7	5		
30-	5	8		
35-		7	4	
40-		6	3	
45-		4	9	
50-		2	7	5
55-60			10	15

أوجد معامل الإرتباط بين عمر الأم وعدد الأطفال .

أسئلة تقويم ذاتي



1. متي نستخدم المقاييس الإحصائية التالية:
   - معامل الاقتران.
   - معامل التوافق
2. عرف الإنحدار الخطي البسيط
3. أكتب معادلة خط الإنحدار (X,Y) .
4. أذكر الطرق التي يتم بها رسم خط الإنحدار.
5. أكتب معادلة أي من الطرق التي يتم بها رسم خط الإنحدار.

الخلاصة

عزيزي الدارس،

ما الذي ناقشناه في الوحدة السادسة

- تعرضنا في الوحدة السادسة إلى الإرتباط والإنحدار الخطي. من خلال المقاييس الإحصائية الستة الواردة فيها.
- تهدف هذه المقاييس إلى دراسة أساليب قياس مقدار واتجاه العلاقة بين متغيرين وقوة الإرتباط بينهما ومن ثم تحديد علاقة رياضية تربط بين المتغيرين وتمكن من التنبؤ بقيمة أحد المتغيرين إذ علم قيمة المتغير الآخر.
- يوجد اختلاف بين العلاقة الدالية والعلاقة الإحصائية.
- توجد أنواع مختلفة من أشكال الإرتباط وقوته وقياسه رقمياً. كما توجد عدة طرق رسم خط الإنحدار.
- هنالك علاقة بين معامل الإرتباط ومعامل الإنحدار.

أرجو أن تراجع هذه الخلاصة ما ورد في هذه الوحدة. حتى توفق في فهم هذه الوحدة والاستفادة منها.

لمحة مسبقة عن الوحدة التالية:

تتحدث الوحدة ا لتالية عن الأرقام القياسية من خلال تعريف الرقم القياسي ومجالات تطبيقها.

تتعرض في ذلك إلى مناسيب الأسعار وخصائصها وأواعها كما تبين سلاسل المناسيب والمشكلات التطبيقية التي تواجه استعمال هذه السلاسل.

كما تشمل طرق حساب الأرقام القياسية مبينة مزايا وعيوب كل طريقة.

أتمني لك التوفيق في دراستك لها.