الفصل الثامن

الارتباط و الانحدار الخطي

لاحظنا في الفصول السابقة الطرق والاساليب المختلفة في جمع وتصنيف وتبويب البيانات كذلك عملية استخراج بعض المقاييس التي تعطي فكرة اكثر وضوحاً عن تلك البيانات كالمتوسطات ومقاييس التشتت والعزوم ومقاييس الالتواء . ان هذه الطرق والاساليب استندت على البيانات المجمعة عن متغير واحد فقط سواء كانت هذه البيانات مبوبة في توزيع تكراري ام غير ذلك . وفي احوال كثيرة يواجه الباحث حالات تتطلب دراسة متغيرين او اكثر في آن واحد لبيان طبيعة ونوع العلاقة التي ترتبط بها هذه المتغيرات . عليه فان هذا الفصل سوف يخصص لدراسة مقاييس اخرى تحدد درجة ونوع وشكل العلاقة بين متغيرين او اكثر .

٨ ـ ١ : الارتباط الخطي **Linear correlation**

ان مفهوم الارتباط الخطي يقترن بحالة وجود متغيرين او اكثر ترتبط مع بعض بعلاقات خطية معينة على سبيل المثال العلاقة بين طول الشخص (سم) ووزنه (كغم) ، العلاقة بين تحصيل الطالب المتخرج من الكلية ومعدل درجاته في الثانوية والمستوى المعاشي لاسرته ، العلاقة بين نسبة الشفاء من مرض معين وكمية الجرعة من الدواء المخصص للمريض وعمر المريض ، وغيرها من الامثلة الاخرى ، فاذا كان التغير في احد المتغيرات يؤثر في تغير متغير آخر او مجموعة متغيرات اخرى عندئذ يقال ان هذه المتغيرات مرتبطة فيما بينها . واذا كان المتغيرين المرتبطين (او مجموعة المتغيرات المرتبطة) يتغيران (تتغير) بنفس الاتجاه اي زيادة (نقصان) في احدهما تؤدي الى زيادة (نقصان) في الاخر (الاخرى) عندئذ يقال ان الارتباط مابينهما هو ارتباط موجب . على سبيل المثال زيادة طول الشخص يتوقع ان يقابلها زيادة في وزنه . انخفاض في دخل الفرد يتوقع منه انخفاض في انفاقه على بعض السلع . اما اذا كان المتغيرين المرتبطين (او مجموعة من المتغيرات المرتبطة)

في احدهما تؤدي الى نقصان (نقصان) في احدهما تؤدي الى نقصان يتغيراته (تتغير) باتجاه معاكس اي زيادة (نقصان) عندئذ يقال ان الارتباط مابينهما هو ارتباط (زيادة) في الاخر (الاخرى) عندئذ يقال ان الارتباط مابينهما هو ارتباط سالب على سبيل المثال زيادة في سعر الوحدة من سلعة معينة يتوقع أن يؤدي الى انخفاض في الطلب على تلك السلعة، انخفاض في درجات الحرارة يتوقع ان ينجم عنه زيادة الطلب على الوقود (كالبترول مثلاً). ويقال ان الارتباط بين متغيرين او اكثر هو ارتباط تام perfect اذا كان التغير في احدهما متناسب مع التغير في المتغير الاخر (الاخرى) على سبيل المثال ان الارتباط بين درجة الحرارة المئوية ودرجة الحرارة الفهرنهايتية هو ارتباط تام باعتبار ان التغير في الاولى متناسب مع التغير في الثانية وهذا يعني ان احداهما مشتقة من الاخرى هذه العلاقة التي تربط فيا بينهما هي $\left(F = \frac{9}{5} C + 32 \right)$.

ان الاساس في دراسة الارتباط وحساب قيمته (كما سنلاحظ لاحقاً) يستند الى العلاقة السببية التي تربط المتغيرات مع بعض، اي مانعنيه وجود علاقة منطقية تفسر سبب الارتباط بينها. فمثلاً تغير سلوك المستهلك تجاه زيادة استهلاكه من سلعة معينة ناجم عن عامل او جملة عوامل جعلته يغير سلوكه هذا مثلاً زيادة دخله الشهري، زيادة عدد افراد عائلته، ارتفاع اسعار سلع بديلة اخرى لهذه السلعة، تقليص انفاقه على سلع اخرى، ... الخ. لاحظ هنا المنطق في العلاقة التي تربط هذه المتغيرات. في حين لايوجد هناك اي منطق او سبب للعلاقة مابين زيادة عدد حوادث الطرق من جهة وارتفاع استهلاك الفرد كمعدل من الطاقة الكهربائية على الرغم من امكانية حساب قيمة الارتباط فيا بينهما الا انها قيمة ليس لها اي معنى او تفسير.

لذلك وقبل القيام بعملية حساب الارتباط يستوجب الامر وجود علاقة منطقية تربط المتغيرات وتفسر السبب الحاصل في تغيرها.

ويتم حساب الارتباط من خلال معامل يدعى بمعامل الارتباط correlation coefficient. وعليه يعرف معامل الارتباط بأنه درجة او قيمة العلاقة التي تربط متغيرين او اكثر مع بعض. وهي قيمة حقيقية خالية من وحدات قياس المتغيرات المرتبطة بعلاقة.

والارتباط على انواع عديدة كل منها يختص بنوع معين من العلاقات. فهناك ارتباط للعلاقة بين متغيرين، ارتباط بين اكثر من متغيرين، ارتباط بين المتغيرات من النوع الوصفي وغيرها. وفيا يلي عرض لكل نوع من هذه الانواع:

٨ ـ ١ ـ ١ : الارتباط الخطي البسيط Simple linear correlation

يعرف الارتباط الخطي البسيط بانه درجة او القيمة العددية للعلاقة بين متغيرين. وفيما يلي المفاهيم الاساسية التي توضح فكرة الارتباط البسيط اضافة الى طرق حسابه وخصائصه.

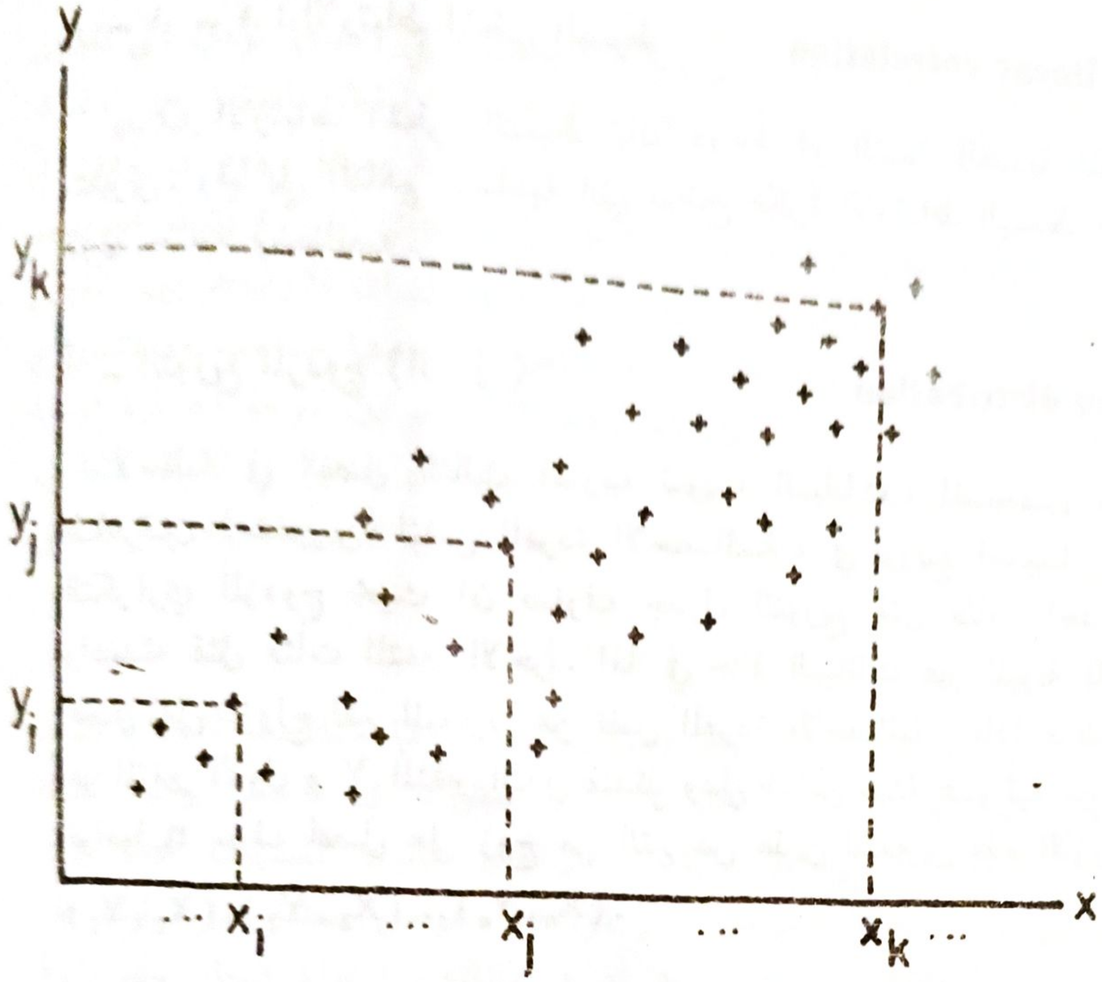
أ ـ التوزيع المزدوج (الثنائي) Bivariate distribution

لاحظنا في الفصل الثالث اسلوب تبويب البيانات، المستحصل عليها عن ظاهرتين (متغيرين) لنفس المفردة الاحصائية، في توزيع اسميناه بالتوزيع التكراري المزدوج بحيث ان صفوف جدول التوزيع تمثل فئات احد المتغيرين واعمدته تمثل فئات المتغير الاخر. اما في حالة البيانات غير المبوبة فاننا سوف نحصل على ازواج لقيم المتغيرين عن نفس المفردة الاحصائية. فاذا فرضنا ان X هو المتغير الاول و Y المتغير الثاني عندئذٍ وعلى اساس عينة عشوائية من المفردات توامها n سوف نحصل على زوج من القيم عن هذين المتغيرين هذه الازواج هي:
$(X_1, Y_1), (X_2, Y_2) \dots (X_n, Y_n)$.

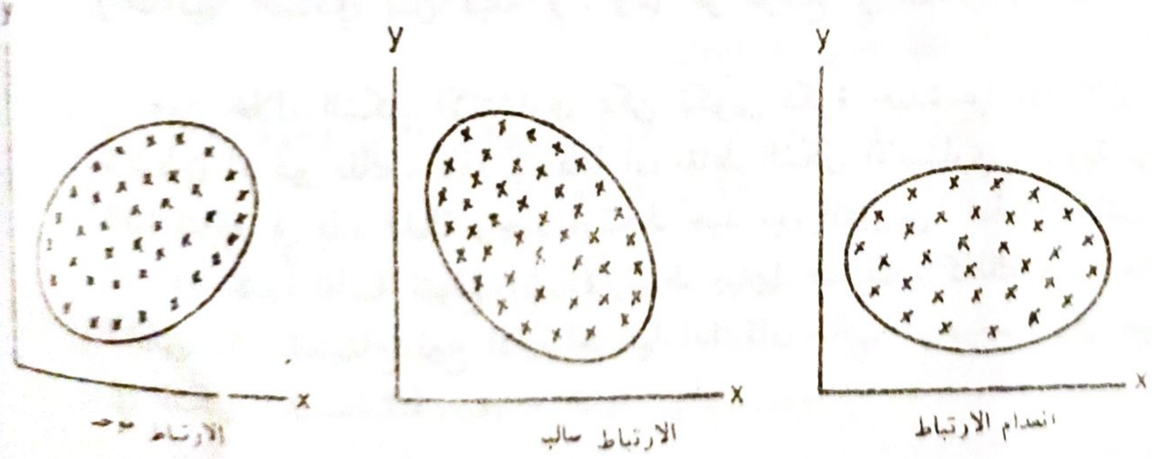
ب ـ الشكل الانتشاري Scatter diagram

ان الشكل الانتشاري يعتبر ابسط طريقة لعرض بيانات توزيع مزدوج. وهو عبارة عن انتشار النقاط في المستوي xy التي احداثيها السيني يمثل قيمة x واحداثيها الصادي يمثل قيمة y. وكما هو موضح في الشكل (٨ ـ ١).

ومن خلال الشكل الانتشاري يمكن تكوين فكرة جيدة عما اذا كان المتغيرين مرتبطين ام غير ذلك. فاذا لاحظنا ان نقاط الشكل الانتشاري متقاربة من بعضها فاننا نتوقع في هذه الحالة وجود ارتباط جيد بين المتغيرين. اما اذا كانت النقاط متباعدة كثيراً فاننا نتوقع ان الارتباط بينهما ضعيف. كذلك ومن خلال هذا الشكل يمكن استنتاج نوع الارتباط فيما اذا كان سالب ام موجب وكما هو موضح في الشكل (٨ ـ ٢):



شكل (٨ ـ ١) : شكل انتشاري



شكل (٨ ـ ٢) : توضيح مفهوم الارتباط البسيط

جـ ـ معامل الارتباط الخطي البسيط Simple correlation coefficient

يعرف معامل الارتباط الخطي البسيط بانه القيمة العددية للعلاقة الخطية بين متغيرين. ويعتبر العالم الانكليزي كارل پيرسون (1867-1936) اول من وضع صيغة لهذا المعامل. وغالباً مايرمز لهذا المعامل بالرمز r_{xy} او r_{12} او r . وفيما يلي طرق حساب هذا المعامل لبيانات غير مبوبة ولبيانات مبوبة.

اولاً : حساب معامل الارتباط البسيط لبيانات غير مبوبة :

لتكن $(X_i, Y_i), i = 1, 2, \ldots, n$ تمثل ازواج القيم المستحصل عليها من عينة من المفردات قوامها n . وافرض ان S_y, S_x يمثلان على التوالي الانحراف المعياري لقيم X ، Y عندئذٍ :

$$r_{xy} = \frac{\operatorname{cov}(X, Y)}{S_x \cdot S_y} = \frac{S_{xy}}{S_x \cdot S_y}$$

حيث ان $\operatorname{cov}(X, Y) = S_{xy}$ يسمى التباين المشترك Covariance بين X, Y ويتم حسابه وفق الصيغة التالية :

$$\operatorname{cov}(X, Y) = S_{xy} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

وعليه فان

$$r_{xy} = \frac{\dfrac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{n}}{\sqrt{\dfrac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}} \cdot \sqrt{\dfrac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n}}}$$

$$= \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \cdot \sum_{i=1}^{n} (Y_i - \bar{Y})^2}} \quad \ldots (1)$$

يلاحظ من الصيغة (١) ان المقام هو كمية موجبة دائماً (لماذا). وهذا يعني ان اشارة معامل الارتباط تتحدد من خلال اشارة البسط. فاذا كان البسط (التباين المشترك) موجباً فذلك يعني ان الارتباط موجب والعكس صحيح ايضاً. كذلك فان $r_{xy} = r_{yx}$.

مثال (١) : البيانات التالية تمثل الكمية المعروضة من سلعة معينة وسعر الوحدة الواحدة من هذه السلعة. يطلب حساب معامل الارتباط البسيط بين الكمية المعروضة والسعر.

الكمية المعروضة (Y) :	3	5	7	8	9	11	6	8	6
السعر (X) :	2	2	5	4	5	6	3	5	4

الحل : ان $\bar{Y} = 7$ وان $\bar{X} = 4$

نعمل الجدول التالي :

$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$X_i - \bar{X}$	$Y_i - \bar{Y}$	X_i	Y_i
8	4	16	− 2	− 4	2	3
4	4	4	− 2	− 2	2	5
0	1	0	1	0	5	7
0	0	1	0	1	4	8
2	1	4	1	2	5	9
8	4	16	2	4	6	11
1	1	1	− 1	− 1	3	6
1	1	1	1	1	5	8
0	0	1	0	− 1	4	6
24	16	44	0	0	36	63

وعليه فان

$$r_{xy} = \frac{24}{\sqrt{16} \cdot \sqrt{44}} = \frac{24}{26.533} = 0.905$$

وهذا يعني ان درجة الارتباط مابين الكمية المعروضة من هذه السلعة وسعر الوحدة منها هو 0.915 وانه ارتباط موجب (دلالة على انه كلما ازداد السعر ازدادت بالمقابل لكمية المعروضة من هذه السلعة) على ضوء هذه البيانات .

ان الصيغة (١) قد تبدو اسلوب مطول لحساب الارتباط وخصوصاً اذا كان عدد مفردات العينة كبير . وهناك صيغ اخرى مشتقة من الصيغة (١) يتم من خلالها التعامل مع قياسات العينة مباشرة دون اللجوء الى حساب الانحرافات . هذه الصيغ هي الاتي :

أن

$$r_{xy} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 . \sum_{i=1}^{n} (Y_i - \bar{Y})^2}}$$

بفتح الاقواس في البسط والمقام واجراء بعض العمليات الجبرية البسيطة نحصل على مايلي :

$$r_{xy} = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sqrt{\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)\left(\sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2\right)}} \quad \ldots (2)$$

او ان

$$r_{xy} = \frac{n\sum_{i=1}^{n} X_i Y_i - \left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{\sqrt{\left\{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2\right\}.\left\{n\sum_{i=1}^{n} Y_i^2 - \left(\sum_{i=1}^{n} Y_i\right)^2\right\}}}$$

مثال (٢) : لبيانات المثال (١) جد معامل الارتباط البسيط باستخدام الصيغتين (2) و (3) .

الحل : نعمل الجدول التالي :

X_iY_i	X_i^2	Y_i^2	X_i	Y_i
6	4	9	2	3
10	4	25	2	5
35	25	49	5	7
32	16	64	4	8
45	25	81	5	9
66	36	121	6	11
18	9	36	3	6
40	25	64	5	8
24	16	36	4	6
276	160	485	36	63

$$\therefore r_{xy} = \frac{276 - 9(4)(7)}{\sqrt{(160 - 9(4)^2).(485 - 9(7)^2)}} \qquad \ldots(2)$$

$$= \frac{(9)(276) - (36)(63)}{\sqrt{\{9(160) - (36)^2\}.\{9(485) - (63)^2\}}} \qquad \ldots(3)$$

$$= \frac{24}{\sqrt{(16)(44)}} = 0{\cdot}905$$

ثانياً : حساب معامل الارتباط البسيط لبيانات مبوبة ..

افرض وجود توزيع تكراري مزدوج عدد صفوفه (فئات المتغير الاول) هو k وعدد اعمدته (فئات المتغير الثاني) هو m . وهذا يعني ان عدد خلايا هذا التوزيع هو kxm . افرض ان f_{ij} , $j = 1, 2, \ldots, m$, $i = 1, 2, \ldots, k$ يمثل تكرار الخلية المقابلة للفئة i للمتغير الاول والفئة j للمتغير الثاني. وافرض ايضاً مايلي :

ـ ان x_i تمثل مراكز فئات المتغير X (الصفوف مثلاً).
ـ ان y_i تمثل مراكز فئات المتغير Y (الاعمدة مثلاً)
ـ ان f_i تمثل مجاميع التكرارات المقابلة لفئات X.
ـ ان f_j تمثل مجاميع التكرارات المقابلة لفئات Y.
ـ ان L_i تمثل طول كل فئة من فئات X.
ـ ان L_j تمثل طول كل فئة من فئات Y.
ـ ان X' يمثل مركز فئة اختياري من بين مراكز فئات X.
ـ ان y' يمثل مركز فئة اختياري من بين مراكز فئات Y.
ـ ان $u_i = (X_i - X')/L_i$

$v_j = (Y_j - Y')/L_j$

ـ ان $u_i f_i$ تمثل حاصل ضرب u_i في مجموع التكرار المقابل للفئة i الى x.

ـ ان $v_j f_j$ تمثل حاصل ضرب v_j في مجموع التكرار المقابل للفئة j الى y.
ـ ان $u_i^2 f_i$ تمثل حاصل ضرب مربع u_i في مجموع التكرار المقابل للفئة i الى x

ـ ان $v_j^2 f_j$ تمثل حاصل ضرب مربع v_j في مجموع التكرار المقابل للفئة j الى y.

ـ ان $u_i v_j f_{ij}$ تمثل حاصل ضرب u_i في v_j في تكرار الخلية المقابلة للفئة i من x والفئة j من y.
ـ n تمثل المجموع الكلي للتكرارات في التوزيع المزدوج.
ـ $\Sigma f_i = \Sigma f_j = n$

عندئذ وفق هذه المعطيات يمكن حساب معامل الارتباط البسيط مابين x و y وفق الصيغة التالية:

$$r_{xy} = \frac{n \sum_{i=1}^{k} \sum_{j=1}^{m} u_i v_j f_{ij} - \left(\sum_{i=1}^{k} u_i f_i \right) \left(\sum_{j=1}^{m} v_j f_j \right)}{\sqrt{\left\{ n \sum_{i=1}^{k} u_i^2 f_i - \left(\sum_{i=1}^{k} u_i f_i \right)^2 \right\} . \left\{ n \sum_{j=1}^{m} v_j^2 f_j - \left(\sum_{j=1}^{m} v_j f_j \right)^2 \right\}}}$$

وكما هو موضح في الجدول التالي:

فئات X / فئات Y	1	2	…	m	f_i	x_i	u_i	u_if_i	$u_i^2f_i$	$u_iv_jf_{ij}$
1	f_{11}	f_{12}	…	f_{1m}	f_1	x_1	u_1	u_1f_1	$u_1^2f_1$	$\sum_{j=1}^{m} v^*_{1j}$
2	f_{21}	f_{22}	…	f_{2m}	f_2	x_2	u_2	u_2f_2	$u_2^2f_2$	$\sum_{j=1}^{m} v^*_{2j}$
⋮	⋮	⋮	… f_{ij} …	f_{im} ⋮	f_i ⋮	⋮	⋮	⋮		⋮
k	f_{k1}	f_{k2}	…	f_{km}	f_k	x_k	u_k	u_kf_k	$u_k^2f_k$	$\sum_{j=1}^{m} v^*_{kj}$
f_j	f_1	f_2	… f_j …	f_m	n	–	–	$\sum u_if_i$	$\sum u_i^2f_i$	v^*
y_j	y_1	y_2	…	y_m	–					
v_j	v_1	v_2	…	v_m	–					
v_jf_j	v_1f_1	v_2f_2	…	v_mf_m	$\sum v_jf_j$					
$v_j^2f_j$	$v_1^2f_1$	$v_2^2f_2$	…	$v_m^2f_m$	$\sum v_j^2f_j$					
$v_ju_if_{ij}$	$\sum_{i=1}^{k} v^*_{i1}$	$\sum_{i=1}^{k} v^*_{i2}$	…	$\sum_{i=1}^{k} v^*_{im}$	v^*					

حيث ان $\sum_{i=1}^{k} v^*_{ij} = \sum_{i=1}^{k} v_j u_i f_{ij} \, , j = 1, 2, \ldots, m$.

وان $\sum_{j=1}^{m} v^*_{ij} = \sum_{j=1}^{m} u_i v_j f_{ij} \, , i = 1, 2, \ldots, k$

وان $v^* = \sum_{j=1}^{m} \sum_{i=1}^{k} v^*_{ij} = \sum_{i=1}^{k} \sum_{j=1}^{m} v^*_{ij}$

مثال : الاتي توزيع تكراري مزدوج لدرجات 100 طالب في مادتي الاحصاء (x) والرياضيات (Y) . يطلب حساب معامل الارتباط البسيط مابين X ، Y .

فئات Y / فئات X	30 –	40 –	50 –	60 –	70 – 80
40 –	2	4	6	–	–
50 –	3	6	7	1	–
60 –	1	3	10	4	1
70 –	–	3	5	8	9
80 –	–	–	2	5	11
90 – 100	–	–	1	3	5

الحل : نعمل الجدول التالي :

Y / X	30 –	40 –	50 –	60 –	70 – 80	f_i (المجموع)	x_i	u_i	$u_i f_i$	$u_i^2 f_i$	$u_i v_j f_{ij}$
40 –	2 (8)	4 (4)	6 (0)	– (0)	– (0)	12	45	– 2	– 24	48	(12)
50 –	3 (6)	6 (6)	7 (0)	1 (– 1)	– (0)	17	55	– 1	– 17	17	(11)
60 –	1 (0)	3 (0)	10 (0)	4 (0)	1 (0)	19	65	0	0	0	(0)
70 –	– (0)	3 (– 3)	5 (0)	8 (8)	9 (18)	25	75	1	25	25	(23)

(54)	72	36	2	85	18	11 (44)	5 (10)	2 (0)	– (0)	– (0)	80 –
(39)	81	27	3	95	9	5 (30)	3 (9)	1 (0)	– (0)	– (0)	90 – 100
139	243	47	–	–	100	26	21	31	16	6	المجموع f_j
					–	75	65	55	45	35	y_j
					–	2	1	0	– 1	– 2	v_j
					45	52	21	0	– 16	– 12	$v_j f_j$
					165	104	21	0	16	24	$v_j^2 f_j$
					139	(92)	(26)	(0)	(7)	(14)	$v_j u_i f_{ij}$

مع ملاحظة انه تم اختيار مركز الفئة الثالثة الى x(65) كمركز فئة اختياري ومركز الفئة الثالثة الى Y (55) كمركز فئة اختياري .

وعليه فان :

$$r_{xy} = \frac{(100)(139) - (47)(45)}{\sqrt{\{(100)(243) - (47)^2\}\{(100)(165) - (45)^2\}}}$$

$$= \frac{11785}{\sqrt{(22091)(14475)}} = 0{\cdot}659$$

وعليه فان درجة الارتباط بين موضوعي الاحصاء والرياضيات على اساس درجات الطلبة في الامتحانين هي 0.659 .

د ـ خصائص معامل الارتباط البسيط

١ ـ ان قيمة معامل الارتباط البسيط تتراوح مابين(1 + ,1 −)

البرهان :

لدينا

$$r = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\Sigma(X_i - \bar{X})^2 . \Sigma(Y_i - \bar{Y})^2}}$$

افرض ان

$$a_i = X_i - \bar{X},\ b_i = Y_i - \bar{Y},\ i = 1, 2, \ldots, n$$

$$\therefore r = \frac{\Sigma a_i b_i}{\sqrt{(\Sigma a_i^2)(\Sigma b_i^2)}} \rightarrow r^2 = \frac{(\Sigma a_i b_i)^2}{(\Sigma a_i^2)(\Sigma b_i^2)}$$

لكن لاي عددين حقيقين

مثل a, b

$$(\Sigma a_i b_i)^2 \leq (\Sigma a_i^2)(\Sigma b_i^2)$$

وذلك حسب متباينة شوارتز Schwartz inquality

فاذن

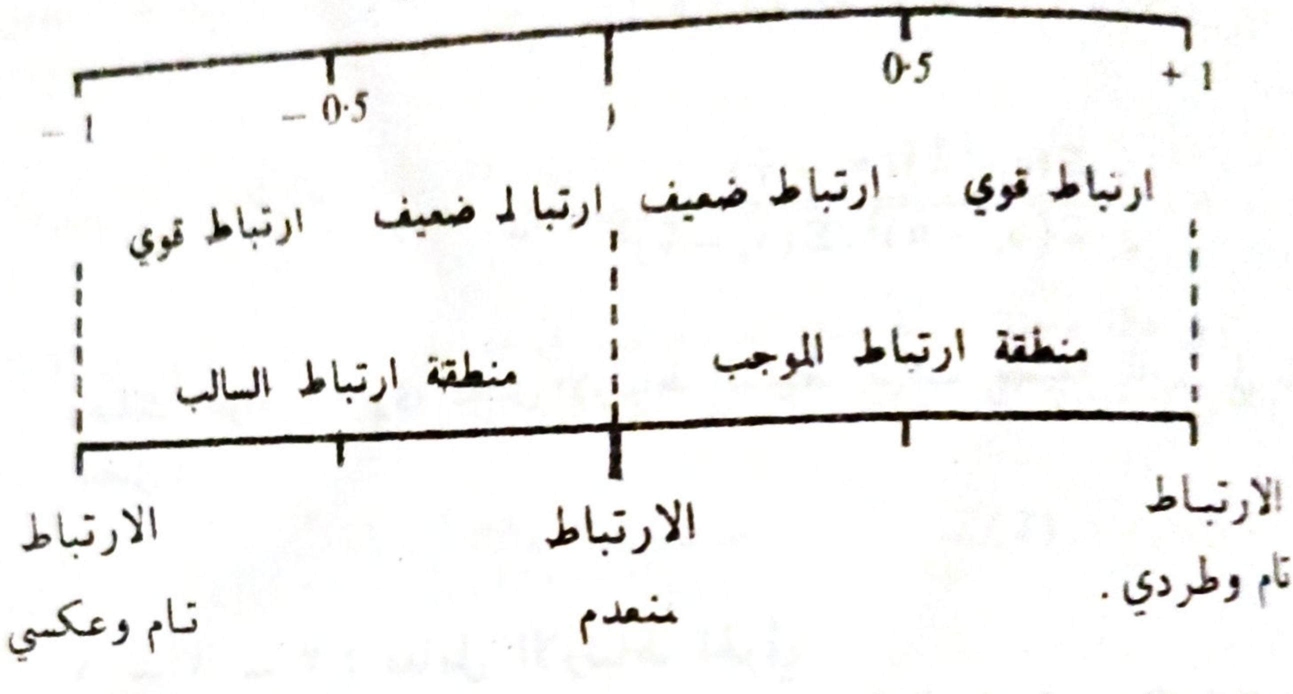
$$r^2 . (\Sigma a_i^2)(\Sigma b_i^2) \leq (\Sigma a_i^2)(\Sigma b_i^2)$$

وعندئذ فان $r^2 \leq 1$ طالما ان $\Sigma a_i^2, \Sigma b_i^2 > 0$

عليه فان

$$-1 \leq r \leq +1$$

والشكل (٨ – ٣) يمثل توضيح لقيم معامل الارتباط الممكنة



شكل (٨ – ٣) : قيم معامل الارتباط الممكنة

٢ – ان اية تحويلات خطية على قيم المتغيرات x,y لاتؤثر على قيمة r .

البرهان :

لتكن $u_i = \frac{X_i - a}{h}$, $v_i = \frac{Y_i - b}{k}$ تمثل تحويلات خطية على قيم المتغيرين x,y حيث ان a, b, h, k تمثل ثوابت اختيارية حقيقية فاذن :

$$X_i = a + hu_i \,,\; Y_i = b + kv_i \,,\; i = 1, 2, \ldots, n$$

عليه فان :

$$\bar{X} = a + h\bar{u} \,,\; \bar{Y} = b + k\bar{v}$$

وان :

$$X_i - \bar{X} = h(u_i - \bar{u}) \,,\; Y_i - \bar{Y} = k(v_i - \bar{v})$$

لدينا

$$r_{xy} = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\Sigma(X_i - \bar{X})^2 . \Sigma(Y_i - \bar{Y})^2}}$$

$$= \frac{\Sigma h(u_i - \bar{u}) . k(v_i - \bar{v})}{\sqrt{\{\Sigma h^2(u_i - \bar{u})^2\}\{\Sigma k^2(v_i - \bar{v})^2\}}}$$

$$= \frac{\Sigma(u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\Sigma(u_i - \bar{u})^2 . \Sigma(v_i - \bar{v})^2}} = r_{uv}$$

وهنالك خواص اخرى لمعامل الارتباط البسيط تركت بصيغة تمارين في نهاية الفصل :

٨ ـ ١ ـ ٢ : معامل الارتباط الجزئي
Partial Correlation Coefficient

يلاحظ في بعض الاحيان وجود ارتباط بين متغيرين يعزى جزئياً الى ارتباط متغير ثالث مرتبط مع كلا المتغيرين . لنفرض ان المتغيرين X_1, X_2 مرتبطين وهنالك متغير آخر مثل X_3 مرتبط مع كلا المتغيرين X_1, X_2 . واننا نرغب في قياس درجة الارتباط بين X_1, X_2 باستبعاد اثر X_3 على كلا المتغيرين . ان المعامل الذي يقيس درجة ارتباط متغيرين باستبعاد اثر متغير ثالث يدعى بمعامل الارتباط الجزئي . على سبيل المثال الارتباط مابين طول الفرد (X_1) وزنه (X_2) يتأثر بارتباط عمر الفرد (X_3) مع كل من X_2, X_1 . كذلك الارتباط بين دخل الاسرة الشهري (X_1) وانفاقها الشهري (X_2) يتأثر بارتباط عدد افراد الاسرة (X_3) مع كل من X_2, X_1 . وربما يكون هنالك اكثر من متغير يرتبط مع المتغيرين X_2, X_1 على سبيل المثال الارتباط بين عدد السكائر المدخنة (X_1) يومياً والاصابة بنوع معين من امراض الرئة (X_2) يتأثر بارتباط عمر المدخن (X_3) وعدد سنوات التدخين (X_4) مع كل من X_2, X_1 .

ويمكن ايجاد قيمة معامل الارتباط الجزئي وفق مايلي

افرض ان X_1, X_2, X_3 تمثل ثلاث متغيرات عشوائية بحيث ان المتغيرين X_1, X_2 مرتبطين وان X_3 مرتبط مع كل من X_2, X_1 . وافرض انه على

اساس عينة عشوائية من المفردات قوامها n تم الحصول على القياسات المتناظرة لهذه المتغيرات ، اي $(X_{1i}, X_{2i}, X_{3i}), i = 1,2,...,n$ ، وان معامل الارتباط البسيط بين اي متغيرين منها هو r_{12}, r_{13}, r_{23} . عندئذ فان معامل الارتباط الجزئي بين X_1, X_2 باستبعاد اثر X_3 هو :

$$r_{12\cdot3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} \quad ...(1)$$

وفي حالة وجود اربعة متغيرات ونرغب في حساب الارتباط الجزئي بين الاول والثاني باستبعاد اثر الثالث والرابع فان هذا المعامل هو :

$$r_{12\cdot34} = \frac{r_{12\cdot3} - r_{14\cdot3} \cdot r_{24\cdot3}}{\sqrt{(1 - r_{14\cdot3}^2)(1 - r_{24\cdot3}^2)}} \quad ...(2)$$

حيث ان $r_{12\cdot3}, r_{14\cdot3}, r_{24\cdot3}$ تمثل معاملات الارتباط الجزئية بين $X_1, X_2 ; X_1, X_4 ; X_2, X_4$ باستبعاد اثر X_3 والتي يمكن الحصول عليها بعد تطبيق الصيغة (١) من خلال اعادة ترميزها حسب تسلسل المتغيرات .

مثال (١) : افرض ان X_1 يمثل دخل الاسرة الشهري وان X_2 يمثل انفاق الاسرة الشهري وان X_3 يمثل عدد افراد الاسرة . افرض وعلى اساس عينة من الاسر قوامها 20 اسرة تم الحصول على مايلي :

$$r_{12} = 0{\cdot}91 ,\ r_{13} = 0{\cdot}39 ,\ r_{23} = 0{\cdot}62$$

جد معامل الارتباط بين دخل الاسرة وانفاقها باستبعاد اثر عدد افراد الاسرة .

الحل : ان المطلوب في السؤال ايجاد $r_{12\cdot3}$ وعليه فان :

$$r_{12\cdot3} = \frac{0{\cdot}91 - (0{\cdot}39)(0{\cdot}62)}{\sqrt{(1 - (0{\cdot}39)^2)(1 - (0{\cdot}62)^2)}} = \frac{0{\cdot}6682}{0{\cdot}7225}$$

$$= 0{\cdot}925$$

مثال (٢) : اذا علمت ان $r_{12} = 0.7 , r_{13} = 0.6 , r_{14} = 0.5 , r_{23} = 0.2$
$r_{24} = 0.3 , r_{34} = 0.4$

جد معامل الارتباط بين المتغيرين الاول والثاني بعد استبعاد اثر المتغيرين الثالث والرابع .

الحل : ان المطلوب في السؤال ايجاد $r_{12\cdot34}$ وعليه فان :

$$r_{12\cdot3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} = \frac{0.7 - (0.6)(0.2)}{\sqrt{(1 - (0.6)^2)(1 - (0.2)^2)}}$$

$$= 0.74$$

$$r_{14\cdot3} = \frac{r_{14} - r_{13} \cdot r_{43}}{\sqrt{(1 - r_{13}^2)(1 - r_{43}^2)}} = \frac{0.5 - (0.6)(0.4)}{\sqrt{(1 - (0.6)^2)(1 - (0.4)^2)}}$$

$$= 0.35$$

$$r_{24\cdot3} = \frac{r_{24} - r_{23} \cdot r_{43}}{\sqrt{(1 - r_{23}^2)(1 - r_{43}^2)}} = \frac{0.3 - (0.2)(0.4)}{\sqrt{(1 - (0.2)^2)(1 - (0.4)^2)}}$$

$$= 0.27$$

بتطبيق الصيغة (2) :

$$\therefore r_{12\cdot34} = \frac{0.74 - (0.35)(0.27)}{\sqrt{(1 - (0.35)^2)(1 - (0.27)^2)}} = 0.72$$

ويلاحظ من الصيغة (1) انه اذا كان $r_{12\cdot3} = 0$ فذلك يعني ان $r_{12} = r_{13} \cdot r_{23}$ وهذا يعني ان $r_{12} \neq 0$ فيما اذا كان X_3 مرتبط مع كل من X_1, X_2 . وعلى الرغم من انه قد يحصل ان يكون X_1, X_2 غير مرتبطين عند استبعاد اثر X_3، الا اننا نلاحظ وجود ارتباط فيما بينهما على نحو غير مباشر ، هذا الارتباط ناجم عن تأثر كلاهما بالمتغير X_3 . كذلك اذا كان $r_{13} = r_{23} = 0$

فذلك يعني ان $r_{12\cdot3} = r_{12}$ اي لايوجد اي اثر للمتغير X_3 على المتغيرين X_2, X_1 . ان قيمة معامل الارتباط الجزئي هي الاخرى تتراوح مابين $(-1, +1)$.

مثال (٣) : اذا علمت ان $r_{12} = r_{13} = r_{23} = r$ عندئذ فان

$$r_{12\cdot3} = r_{13\cdot2}\ r_{23\cdot1} = r/(1+r)$$

الحل :

$$r_{12\cdot3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} = \frac{r - r \cdot r}{\sqrt{(1 - r^2)(1 - r^2)}}$$

$$= \frac{r(1 - r)}{1 - r^2} = \frac{r}{1 + r}$$

وعلى النحو نفسه يمكن بيان ان

$$r_{13\cdot2} = r_{23\cdot1} = \frac{r}{1 + r}$$

٨ - ١ - ٣ : معامل الارتباط المتعدد
Multiple Correlation Coefficient

يلاحظ في كثير من الاحيان ان التغير الذي يطرأ على ظاهرة معينة ناجم عن تغير مجموعة ظواهر اخرى مجتمعة وليس ظاهرة واحدة فقط . على سبيل المثال التغير في انفاق الاسرة الشهري يتأثر بتغير دخلها الشهري وعدد افرادها . التغير في انتاجية الدونم الواحد من الحنطة يتأثر بنوع البذور المستخدمة ، نوع السماد المستخدم ، نوع التربة . الاصابة باحد امراض الرئة يتأثر بعمر المدمن على السكاير ، عدد السكائر المدخنة يومياً ، عدد سنوات التدخين وغيرها من الامثلة . ان المعامل الذي يبين درجة العلاقة بين متغير معين ومجموعة متغيرات اخرى يسمى معامل الارتباط المتعدد .

وعلى افتراض ان X_1, X_2, X_3 ثلاث متغيرات عشوائية ونرغب في قياس درجة العلاقة بين X_1 مع X_2, X_3 فان ذلك يتم وفق الصيغة التالية :

$$R_{1\cdot23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}\cdot r_{13}\cdot r_{23}}{1 - r_{23}^2}} \quad r_{23} \neq \mp 1 \quad \dots(1)$$

وهنالك صيغة أخرى مشتقة من (١) بدلالة معاملات الارتباط الجزئية . هذه الصيغة هي :

$$R_{1\cdot23} = \sqrt{1 - (1 - r_{12}^2)(1 - r_{13\cdot2}^2)} \quad \dots(2)$$

البرهان : بتربيع الصيغة (١) نحصل على

$$R_{1\cdot23}^2 = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}\cdot r_{13}\cdot r_{23}}{1 - r_{23}^2}$$

$$\frac{r_{12}^2(1 - r_{23}^2) + (r_{13} - r_{12}\, r_{23})^2}{1 - r_{23}^2}$$

$$= r_{12}^2 + \frac{(r_{13} - r_{12}\cdot r_{23})^2}{1 - r_{23}^2} = r_{12}^2 + \frac{r_{13\cdot2}^2(1 - r_{12}^2)(1 - r_{23}^2)}{1 - r_{23}^2}$$

$$= r_{12}^2 + r_{13\cdot2}^2(1 - r_{12}^2) = 1 - (1 - r_{12}^2) + r_{13\cdot2}^2(1 - r_{12}^2)$$

$$= 1 - (1 - r_{12}^2)(1 - r_{13\cdot2}^2) \therefore R_{1\cdot23} = \sqrt{1 - (1 - r_{12}^2)(1 - r_{13\cdot2}^2)}$$

ويمكن تعميم الصيغة (2) لاكثر من ثلاث متغيرات . فمثلاً في حالة وجود اربعة متغيرات فان :

$$R_{1\cdot234} = \sqrt{1 - (1 - r_{12}^2)(1 - r_{13\cdot2}^2)(1 - r_{14\cdot23}^2)} \quad \dots(3)$$

مثال (١) : افرض ان X_1 يمثل الانفاق الشهري للاسرة و X_2 يمثل دخلها الشهري و X_3 يمثل عدد افراد الاسرة . وعلى اساس عينة عشوائية من الاسر قوامها 20 اسرة تم الحصول على مايلي :

$$r_{12} = 0{\cdot}91,\ r_{13} = 0{\cdot}39,\ r_{23} = 0{\cdot}62$$

يطلب حساب معامل الارتباط المتعدد بين X_1 مع X_2, X_3 .

الحل : باستخدام الصيغة (1) :

$$R_{1.23} = \sqrt{\frac{(0.91)^2 + (0.39)^2 - 2(0.91)(0.39)(0.62)}{1 - (0.62)^2}}$$

$$= \sqrt{\frac{0.5401}{0.6156}} = 0.937$$

الحل : باستخدام الصيغة (2) :

$$r_{13.2} = \frac{r_{13} - r_{12} \cdot r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}}$$

$$= \frac{0.39 - (0.91)(0.62)}{\sqrt{(1 - (0.91)^2)\cdot(1 - (0.62)^2)}}$$

$$= -0.536$$

$$\therefore R_{1.23} = \sqrt{1 - (1 - (0.91)^2)(1 - (-0.536)^2)}$$

$$= 0.937$$

مثال (٣) : اذا علمت ان $r_{12} = r_{13} = r_{23} = r$ عندئذ فان

$$R_{1.23} = R_{2.13} = R_{3.12} = \sqrt{\frac{2r^2}{1 + r}}$$

الحل :

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12} \cdot r_{13} \cdot r_{23}}{1 - r_{23}^2}}$$

$$= \sqrt{\frac{2r^2 - 2r^3}{1 - r^2}} = \sqrt{\frac{2r^2(1 - r)}{(1 - r)(1 + r)}} = \sqrt{\frac{2r^2}{1 + r}}$$

ان مفهوم الارتباط المتعدد يقترن بموضوع الانحدار الخطي المتعدد الذي سنأتي الى ذكره في الفقرات اللاحقه حيث ان مربع معامل الارتباط المتعدد R^2 يسمى بمعامل التحديد Coefficient of determination . وفيما يلي اهم خصائص معامل الارتباط المتعدد :

١ ـ ان قيمته تتراوح مابين الصفر والواحد . اي ان $0 \leq R \leq 1$.

٢ ـ ان قيمته لاتقل عن قيمة اي معامل ارتباط بسيط بين اي متغيرين دخلت في حساب قيمته . اي ان $R_{1\cdot23} \geq r_{12}, r_{13}, r_{23}$.

٣ ـ اذا كانت قيمته مساوية للصفر فذلك يعني ان كافة معاملات الارتباط البسيطة والجزئية التي يدخل المتغير X_1 في حسابها مساوية للصفر .

٤ ـ ان قيمته تتزايد بزيادة عدد المتغيرات التي تؤثر في المتغير X_1 . وهذا يعني ان R دالة غير متناقصة بزيادة عدد المتغيرات المؤثرة في X_1

اي انه اذا كانت X_5, X_4, X_3, X_2 متغيرات تؤثر على X_1 فان

$$R_{1\cdot23} \leq R_{1\cdot234} \leq R_{1\cdot2345}$$

٨ ـ ١ ـ ٤ : ارتباط الرتب Rank Correlation

ان الصيغ السابقة الخاصة لحساب معامل الارتباط (البسيط ، الجزئي ، المتعدد) تستند في الحقيقة على اعتبار ان المتغيرات المعتمدة في الحساب هي متغيرات من النوع الكمي (اي متغيرات ممكنة القياس بوحدات قياس مألوفة كمقياس الطول ، الوزن ، العمر ... الخ) . الا انه ومن الناحية العملية هنالك الكثير من الحالات التي تكون فيها المتغيرات من النوع الوصفي (اي متغيرات غير قابلة للقياس كالمهنة ، الحالة الاجتماعية ، تقديرات درجات ، وغيرها) وعندئذٍ وبهدف حساب الارتباط بين متغيرات من هذا النوع فانه لايمكن استخدام الصيغ السابقة مباشرة لهذا الغرض بل اجراء بعض التحويرات عليها بالشكل الذي يجعلها ممكنة الاستخـــدام لهــذه الاحوال . إن المعـــامـــل الـــذي يقيس درجـــة الارتباط مابين صفتين يدعى بمعامل ارتباط الرتب لسبيرمان Spearman's rank Correlation Coeff. وهنالك معامل اخر هو معامل ارتباط الرتب لكندال . الا اننا سنكتفي هنا بذكر المعامل الاول نظراً لاهميته .