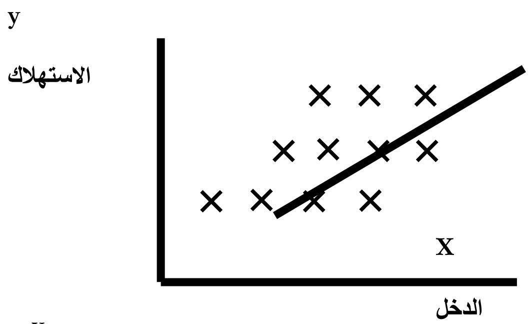
ثالثا: الارتباط والانحدار الخطي

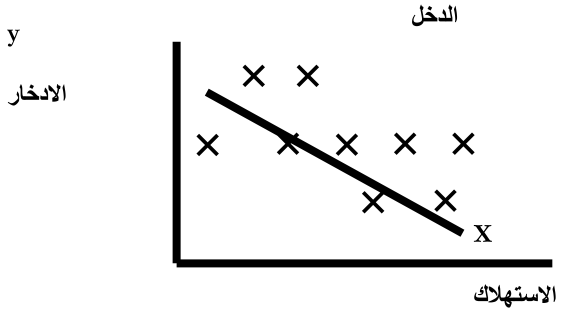
- تعريف الارتباط:

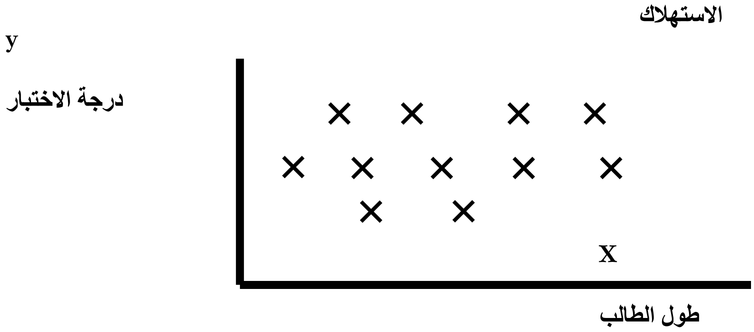
هو علاقة بين ظاهرتين أو متغيرين (x,y) بمعنى أنه إذا تغير أحد المتغيرين فإن الآخر يتبعه .إما في نفس الاتجاه فيكون الارتباط طردي، مثل العلاقة بين الدخل والاستهلاك، أو في اتجاه مضاد فيكون الارتباط عكسي، مثل العلاقة بين الاستهلاك والادخار.

وإذا كانت الظاهرتين مستقلتين يكون الارتباط منعدم، مثل العلاقة بين طول الطالب ودرجة الاختبار.

وشكل الانتشار يوضح نوع هذه العلاقة (طردية – عكسية – منعدمة)







2- معامل ارتباط بيرسون : (الخطي)

هو معامل رقمي يوضح نوع ودرجة العلاقة بين ظاهرتين أو متغيرين، ويرمز له بالرمز (r).

فبفرض أن لدينا مجموعة من(n) مفردة من طلاب إحدى الجامعات وحصلنا منها على بيانات عن أطوالهم (x) وأوزانهم (y)، فتكون البيانات التي لدينا على الصورة:

x الطول : $x_1 , x_2, x_3, ...,x_n$

y الوزن : $y_1,y_2,y_3,........,y_n$

ويحسب معامل الارتباط بين الظاهرتين x,y من العلاقة الآتية :

$$r = \frac{\frac{\sum xy}{n} - \bar{x}\bar{y}}{s_x \;.\; s_y} \qquad (1)$$

حيث : قيم الظاهرتين : x , y

، الوسط الحسابي للظاهرة x : $\overline{X}$ $\qquad \bar{x} = \frac{\sum x}{n}$

، الوسط الحسابي للظاهرة y : $\bar{y}$ $\qquad \bar{y} = \frac{\sum y}{n}$

S_x ، الانحراف المعياري للظاهرة x : $\qquad s_x = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$

S_y ، الانحراف المعياري للظاهرة y : $\qquad s_y = \sqrt{\frac{\sum y^2}{n} - \bar{y}^2}$

عدد المفردات (عدد أزواج القيم) : n

3- خصائص معامل الارتباط :

- قيمة r تكون موجبة في حالة الارتباط الطردي وتكون سالبة في حالة الارتباط العكسي، وتساوي صفر في حالة الارتباط المنعدم .
- قيمة r تساوي 1+ في حالة الارتباط الطردي التام ، 1- في حالة الارتباط العكسي التام .
- قيمة r تتراوح بين 1± وتزداد قوتها كلما قربت من الواحد الصحيح.

مثال (1): لدراسة العلاقة بين الدخل (x) والاستهلاك (y) بآلاف الدولارات، كانت لدينا النتائج الآتية:

$\sum x = 120 \qquad \sum y = 100 \qquad \sum xy = 516$

$\sum x^2 = 720 \qquad \sum y^2 = 410 \qquad n = 40$

احسب معامل ارتباط بيرسون بين الظاهرتين .

الحل:

$$r = \frac{\frac{\sum xy}{n} - \bar{x}\,\bar{y}}{s_x \quad . \quad s_y}$$

$$\bar{x} = \frac{\sum x}{n} = \frac{120}{40} = 3$$

$$s_x = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2} = \sqrt{\frac{720}{40} - (3)^2} = \sqrt{18 - 9} = \sqrt{9}$$

$$\bar{y} = \frac{\sum y}{n} = \frac{100}{40} = 2.5$$

$$s_y = \sqrt{\frac{\sum y^2}{n} - \bar{y}^2} = \sqrt{\frac{410}{40} - (2.5)^2}$$

$$= \sqrt{10.25 - 6.25} = \sqrt{4}$$

$$r = \frac{\frac{516}{40} - (3)(2.5)}{\sqrt{9}\ \ \sqrt{4}} = \frac{12.9 - 7.5}{\sqrt{36}} = \frac{5.4}{6}$$

الارتباط طردي قوي . $\therefore r = 0.9$

مثال (2): الجدول الآتي يوضح العلاقة بين الدخل (x) بمئات الدولارات والاستهلاك (y) بمئات الريالات لعينة من الأسر :

الدخــل (x)	5	4	5	6	9	10	9	12	11	9
الاستهلاك (y)	5	4	5	5	8	6	8	11	10	8

والمطلوب حساب معامل ارتباط بيرسون بين الظاهرتين .

4- معامل ارتباط الرتب (سبيرمان) :

يستخدم في حالة البيانات الوصفية وكذلك الأعداد الكبيرة ، وهو يعطي قيمة تقريبية لمعامل الارتباط بين الظاهرتين. ونحصل عليه بترتيب قيم كل ظاهرة ترتيبا تصاعديا (أو تنازلياً) ثم نحسب معامل الارتباط بين رتب (ترتيب) الظاهرتين بدلا من قيمهما ، باستخدام القانون الآتي :

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

حيث: الفرق بين رتب (ترتيب) x ، ورتب (ترتيب) y : d

مثال: احسب معامل ارتباط الرتب (سبيرمان) بين تقديرات الطلاب في مادتي الإحصاء والرياضيات من البيانات الآتية :

تقديرات الإحصاء (x)	D	C	B	C	D	C	A	F
تقديرات الرياضيات (y)	F	D	A	D	F	B	C	D

5- الانحدار الخطي البسيط:

إذا كان لدينا متغيران، أولهما هو المتغير x ويتم تحديد قيمه مسبقاً بواسطة الباحث، ولتكن هذه القيم هي: $x_1, x_2, x_3, \ldots, x_n$ ويسمى المتغير x بالمتغير المستقل .

ويرافق المتغير x متغير آخر هو y بحيث إذا تغيرت قيمة x تغيرت أيضا قيمة y ، ولنفرض أن قيم y المرافقة لقيم x هي:

$y_1, y_2, y_3, \ldots, y_n$ ويسمى المتغير y بالمتغير التابع .

وفي الانحدار الخطي البسيط، نجد أن المتغير y يعتمد على متغير مستقل واحد هو x ، وأن العلاقة بين المتغيرين (x,y) علاقة خطية، وبالتالي فإن العلاقة التي تربط بين المتغيرين (x,y) هي من النوع :

$$y_i = \beta_o + \beta_1 x_i + \ell_i \qquad i = 1,2,\ldots\ldots,n$$

حيث : β_0 هي الجزء المقطوع من محور y

β_1 هي ميل الخط المستقيم أو معامل انحدار y/x

وهما يعرفان بمعالم العلاقة الخطية

هو الـ ℓ_i عشوائي في تحديد قيمة y_i

ونظراً لأن المعلمتان $\beta_0 \ , \ \beta_1$ مجهولتين ، فإننا نقدرهما من مشاهدات العينة باستخدام عدة طرق أهمها هي طريقة المربعات الصغرى ، وفيها يتم تقدير قيمتي $\beta_0 \ , \ \beta_1$ بالمقدارين b_1 و b_o من العلاقة الآتية :

$$b_1 = \frac{\frac{\sum xy}{n} - \overline{x}\,\overline{y}}{S_x^2} \quad , \quad b_o = \overline{y} - b_1\overline{x}$$

ويكون خط انحدار y/x هو :

$$y_i = b_o + b_1 x_i \qquad (1)$$

وبعد إيجاد القيمة العددية لكل من $b_o \ , \ b_1$ من العلاقة السابقة ، نعوض بقيمها في المعادلة (1) فنحصل على خط (معادلة) انحدار

$$\hat{y} = .7 + .6x \qquad (2)$$

y/x في صورته الخاصة (أي للمسألة التي لدينا) – معادلة رقم 2- ، وبالتعويض في المعادلة (2) بأي قيمة لـ x نحصل على قيمة y المناظرة لقيمة x ، وهذا هو الانحدار .

مثال: من المثال رقم (1) أوجد :

أ- معادلة (خط) انحدار الاستهلاك على الدخل .

ب- تقدير الاستهلاك عندما يصل الدخل (10000) دولار .

الحل:

أ- معادلة (خط) انحدار الاستهلاك على الدخل (y/x) :

$$y_i = b_o + b_1 x_i \qquad (1)$$

ويتم إيجاد قيم الثوابت $b_o \ , \ b_1$ كما يلي :

$$b_1 = \frac{\frac{\sum xy}{n} - \overline{x}\,\overline{y}}{S_x^2} = \frac{5.4}{9} = 0.6$$

$$b_o = \overline{y} - b_1\overline{x} = 2.5 - (.6)3 = 2.5 - 1.8 = .7$$

خط (معادلة) انحدار الاستهلاك على الدخل ، في صورته الخاصة هو :

$$\hat{y} = .7 + 0.6x \qquad (2)$$

ب- تقدير الاستهلاك عندما يصل الدخل 10000 دولار :

Y = ?　　　x = 10000دولار

بالتعويض في المعادلة (2) بقيمة

$$\therefore \hat{y} = .7 + .6(10) = .7 + 6 = 6.7 = 6700$$ دولار

تمرين:

إذا كان معروفاً أن هناك علاقة بين فترة الإصابة بمرض معين وعدد البكتيريا في العضو المصاب، تم اختيار عينة من 12 مصاب بهذا المرض وتم عد البكتيريا عند دخولهم المستشفى وسجلت أطوال فترات إصابتهم فحصلنا

على النتائج التالية :

عدد البكتيريا بالألف (X)	8	7	4	6	9	7	8	3	9	5	7	5
فترة الإصابة باليوم (Y)	11	10	9	8	12	9	13	7	10	6	11	8

المطلوب إيجاد مايلي:

أ- معامل ارتباط بيرسون بين الظاهرتين .

ب- معامل ارتباط سبيرمان (الرتب) بين الظاهرتين .

ج- خط (معادلة) انحدار طول فترة الإصابة على عدد البكتريا، ثم قدر فترة إصابة مريض لديه بكتريا عددها 5000.