الحل:

x	f	xf	$\lvert x-\bar{x}\rvert$	$\lvert x-\bar{x}\rvert f$
5	6	30	5	30
6	3	18	4	12
8	4	32	2	8
15	4	60	5	20
20	3	60	10	30
Total		200		100

$$\bar{X} = \frac{200}{20} = 10$$

$$M.D = \frac{100}{20} = 5$$

رابعاً: الانحراف المعياري (σ) **Standard Deviation**

أ- في حالة المفردات

تعريف: إذا كانت x1, x2,...,xn مجموعة من المشاهدات وسطها الحسابي $\bar{x}$ فإن الانحراف المعياري لها (يرمز له بالرمزσ) يعطى بالعلاقة

$$\boldsymbol{\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}}}$$

$$\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$$

ويسمى مربع الانحراف المعياري $\sigma 2$ بالتباين. (Variance)

$$\sigma^2 = \frac{\sum (x-\bar{x})^2}{n}$$

مثال:

أوجد الانحراف المعياري والتباين للمفردات 1, 2, 3, 4, 5

$$\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$$

$$\bar{X} = \frac{1+2+3+4+5}{5}$$

$$= 3$$

$$\sigma = \sqrt{\frac{(1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2}{5}}$$

$$\sigma = \sqrt{\frac{4+1+0+1+4}{5}}$$

$$\therefore \sigma = \sqrt{2}$$

$$\sigma^2 = 2$$

مثال:

إذا كانت x1, x2,…,xn مجموعة من المشاهدات فأثبت أن:

$$\sigma = \sqrt{\frac{\sum x^2 - n\bar{x}^2}{n}}$$

$$\sigma = \sqrt{\frac{\sum x_i^2 - 2\bar{x}x_i + \bar{x}^2}{n}}$$

بتوزيع المجموع على الحدود نحصل على

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}^2}{n}}$$

لكن

$$\sum_{i=1}^{n} x_i = n\bar{x}$$

$$\therefore \sigma = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - 2\bar{x}(n\bar{x}) + n\bar{x}^2}{n}}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2}{n}}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n}}$$

وهو المطلوب.

ملاحظة: بينا أن

$$\sigma = \sqrt{\frac{\sum x^2 - n\bar{x}^2}{n}}$$

$$\Rightarrow \sigma = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}$$

$$\Rightarrow \sigma = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2}$$

كذلك يمكننا كتابة الصيغة الأخيرة على الصورة

$$\sigma = \sqrt{\overline{x^2} - (\bar{x})^2}$$

يمكننا تلخيص ما سبق بأن قانون الانحراف المعياري يمكن كتابته على عدة صور منها:

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

$$\sigma = \sqrt{\frac{\sum x^2 - n\bar{x}^2}{n}}$$

$$\sigma = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$\sigma = \sqrt{\bar{x}^2 - (\bar{x})^2}$$

ولكن الأكثر استخداماً الشكلين الأول والثالث.

تمرين:

أوجد الانحراف المعياري للمفردات 5 ,4 ,3 ,2 ,1 باستخدام الشكل الثاني لقانون الانحراف المعياري؟

مثال:

إذا كان لدينا سبع مشاهدات بحيث أن $\sum_{i=1}^{7} x_i^2 = 140$ ، $\sum_{i=1}^{7} x_i = 28$ فأوجد الانحراف المعياري لهذه المشاهدات؟

الحل:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{7} x_i^2}{7} - \left(\frac{\sum_{i=1}^{7} x_i}{7}\right)^2}$$

$$\sigma = \sqrt{\frac{140}{7} - \left(\frac{28}{7}\right)^2}$$

$$= \sqrt{20 - 16}$$

$$= 2$$

ب – في حالة الجداول التكرارية

سنتعرض لطريقتين في حساب الانحراف المعياري للجـداول التكراريـة إلا وهـما الطريقـة العامـة وطريقة الوسط الفرضي. وكذلك تستخدم الطريقتين في حساب الانحراف المعياري للمفـردات المتكـررة. باستخدام نفس القوانين.

1- الطريقة العامة لحساب الانحراف المعياري General method

تعريف: إذا كانت $x1, x2,…,xn$ مراكز فئات جدول تكراري وكانت تكرارات فئاته $f1 , f2, … , fm$ على الترتيب فإن

$$\sigma = \sqrt{\frac{\sum_{i=1}^{m}(x_i - \bar{x})^2 f_i}{\sum_{i=1}^{m} f_i}}$$

الصيغة الواردة في التعريف السابق للانحراف المعياري تؤدي إلى صعوبة في العمليـات الحـسابية وخاصة إذا كان الوسط الحسابي عدد غير صحيح. لذلك نستخدم الصيغة التالية:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{m} x_i^2 f_i}{\sum_{i=1}^{m} f_i} - \left(\frac{\sum_{i=1}^{m} x_i f_i}{\sum_{i=1}^{m} f_i}\right)^2}$$

تمرين:

أثبت أن الصيغتين السابقتين للانحراف المعياري في حالة الجداول التكرارية متكافئتين؟

يمكن كتابة الصيغة الثانية للانحراف المعياري في حالة الجداول التكرارية بالصورة

$$\sigma = \sqrt{\overline{x^2} - (\bar{x})^2}$$

مثال:

أوجد التباين (Variance) للجدول التكراري التالي؟

Class	10-14	15-19	20-24	25-29	30-34	Total
Frequency	1	2	6	7	4	20

الحل:

$$\sigma^2 = \frac{\sum x^2 f}{\sum f} - \left(\frac{\sum xf}{\sum f}\right)^2$$

Class	frequency	Class-mark(x)	xf	X2	X2f
10-14	1	12	12	144	144
15-19	2	17	34	289	578
20-24	6	22	132	484	2904
25-29	7	27	189	729	5103
30-34	4	32	128	1024	4096
Total	20		495		12825

$$\sigma^2 = \frac{12825}{20} - \left(\frac{495}{20}\right)^2$$

$= 641.25 - 612.5626$

Variance $\left(\sigma^2\right) = 28.6875$

ويمكننا تبسيط الإجراءات الحسابية باستخدام طريقة الوسط الفرضي التالية:

2- طريقة الوسط الفرضي Assumed-mean method

إذا كان لدينا مجموعة من المشاهدات وأضفنا لها "أو طرحنا منها" مقداراً ثابتاً فإن ذلك لا يـؤثر على تباعد القيم عن بعضها البعض. وذلك هـو المبـدأ الـذي نـستخدمه في حـساب الانحـراف المعيـاري بطريقة الوسط الفرضي.

إذا كانت x1, x2,…,xm مراكز فئات جدول تكراري، بحيث أن التكرارات المقابلة للفئات هي f1 , f2, … , fm على الترتيب فإن قانون الانحراف المعياري بطريقة الوسط الفرضي هو

$$\sigma = \sqrt{\frac{\sum_{i=1}^{m} d_i^2 f_i}{\sum_{i=1}^{m} f_i} - \left(\frac{\sum_{i=1}^{m} d_i f_i}{\sum_{i=1}^{m} f_i}\right)^2}$$

حيث $d_i = x_i - A$

A: assumed-mean (الوسط الفرضي)

والوسط الفرضي تختاره مركز فئة متوسطة الموقع في الجدول ويكون تكرارها كبير نوعا ما.

وبشكل أكثر اختصاراً

$$\sigma = \sqrt{\frac{\sum df}{\sum f} - \left(\frac{\sum df}{\sum f}\right)^2}$$

مثال:

أوجد الانحراف المعياري للجدول التكراري التالي بطريقة الوسط الفرضي.

Class	10-19	20-29	30-39	40-49	50-59	Total
Frequency	1	2	3	3	1	10

الحل:

$$\sigma = \sqrt{\frac{\sum d^2 f}{\sum d} - \left(\frac{\sum df}{\sum d}\right)^2}$$

class	**frequency**	**class mark (x)**	**d=x-34.5**	**df**	**d2**	**d2f**
10-19	1	14.5	-20	-20	400	400
20-29	2	24.5	-10	-20	100	200
30-39	3	34.5	0	0	0	0
40-49	3	44.5	10	30	100	300
50-59	1	54.5	20	20	400	400
Total	10			10		1300

لاحظ أننا اعتبرنا A = 34.5

$$\therefore \sigma = \sqrt{\frac{1300}{10} - \left(\frac{10}{10}\right)^2}$$
$$= \sqrt{130 - 1}$$
$$= \sqrt{129}$$
$$= 11.36$$

معامل الاختلاف **Coefficient of Variation**

يعرف معامل الاختلاف بأنه النسبة المئوية بين الانحراف المعياري والوسط الحسابي ويرمز له بالرمز CV، حيث

$$C.V = \frac{\sigma}{\overline{X}} \times 100\,\%$$

مثال:

إذا كان الانحراف المعياري لتوزيع ما (5.33) والوسط الحسابي (24.75) احسب معامل الاختلاف لهذا التوزيع؟

الحل:

$$C.V = \frac{\sigma}{\overline{X}} \times 100\,\%$$
$$= \frac{5.35}{24.75} \times 100\%$$
$$= 21.61\ \%$$

ويستخدم معامل الاختلاف للمقارنة بين مجموعتين فكلما كان معامل الاختلاف أقل كان تجانس القيم أكثر.

مثال:

إذا كان الوسط الحسابي والانحراف المعياري لإنتاج مصنعين لمدة عشرة سنوات كما يلي:

المصنع A: $\overline{X} = 30,8 = 8$

المصنع B: $\overline{X} = 60,8 = 12$

أي المصنعين إنتاجه أفضل.

الحل:

$$C.V_A = \frac{8}{30} \times 100\% = 26.67\%$$

$$C.V_B = \frac{12}{60} \times 100\% = 20\%$$

بما أن معامل الاختلاف للمصنع الثاني أقل من المصنع الأول فإن إنتاج المصنع الثاني أفضل لأنه أكثر تجانساً.

تمرين:

أعد حل المثال السابق باستخدام الطريقة العامة؟

ملاحظة عامة: جميع مقاييس التشتت لا تتأثر بإضافة ثابت حقيقي لجميع المفردات. ولكنها تتأثر بضرب جميع المفردات بثابت حقيقي. (انظر تمرين5).

تمارين

-1 للمفردات التالية: 1, 4, -1, 0, 6

أوجد:

1. المدى range؟
2. الانحراف المتوسط mean-deviation؟
3. الانحراف المعياري standard deviation؟
4. التباين variance؟

-2 أعد حل السؤال السابق للمفردات: 1, 7, 12, 2, 5, 17, 20, 15, 16, 18, 1, 4, 3, 7, 9 1

-3 للجدول التكراري التالي:

Class	-10-(-1)	0-9	10-19	20-29	30-39	40-49	Total
Frequency	5	7	9	8	10	1	40

أوجد:

1. المدىrange ؟
2. نصف المدى الربيعي semi interquartile range؟
3. الانحراف المتوسط Mean deviation؟
4. التباين variance؟

-4 في السؤال السابق أوجد المدى العشري للتوزيع حيث:

المدى العشيري = العشير التاسع – العشير الأول

Decile range = D9-D1

-5 إذا كان لدينا المفردات x1, x2,...,xn، وعدّلت هذه المفردات حسب العلاقـة y= ax+b، حيث a, b عددان حقيقيان. x المفردة قبل التعديل y المفردة بعد التعديل فأثبت أن:

أ- الانحراف المتوسط بعد التعديل = $|a|$ × الانحراف المتوسط قبل التعديل.

ب- الانحراف المعياري بعد التعديل = $|a|$ × الانحراف المعياري قبل التعديل.

6- إذا كان لدينا مجموعة من المفردات، انحرافها المعياري 5، وعدّلت جميع المفردات حسب العلاقة $y= 6-2x$ فأوجد:

أ- الانحراف المعياري بعد التعديل؟

ب- التباين قبل التعديل؟

ج- التباين بعد التعديل وما علاقته بالتباين قبل التعديل؟

7- إذا كان الانحراف المعياري لعشرة مشاهدات هو (5). أوجد مجموع مربعات انحرافات القيم عن وسطها الحسابي؟

8- إذا كان التباين للقيم 1, 5,k ,4- هو (11.5). أوجد الوسط الحسابي وقيمة k ؟ (كم حلا للسؤال؟).

9- إذا كان مجموع مربعات مئة قيمة هو (1500). ومجموع هذه القيم (300)، جد تباين هذه القيم؟

10- إذا كانت انحرافات ستة قيم عن وسطها الحسابي هي 0 ,6- ,7 ,2- ,5 ,4-. جد الانحراف المعياري لهذه القيم؟ وكذلك الانحراف المتوسط؟

11- إذا كان الانحراف المعياري والوسط الحسابي لرواتب مجموعة موظفين في شركة بالدينار (12)، (180) على الترتيب. فإذا قرر مدير الشركة زيادة الرواتب بمقدار عشرة دنانير لكل موظف. احسب الانحراف المعياري والوسط الحسابي للرواتب بعد التعديل؟

12- للجدول التكراري التالي

Class	100-119	120-139	140-159	160-179	180-199	Total
Frequency	17	13	24	15	11	80

احسب:

أ- الوسط الحسابي بطريقة الوسط الفرضي.

ب- الانحراف المعياري بطريقة الوسط الفرضي.

ج- المدى الربيعي.

د- المدى العشيري.

13- إذا كانت x1, x2,...,xn مجموعة من القيم وكان $\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2 = 272$ والانحراف المعياري لهذه القيم (4) جد n؟

14- إذا كان نصف المدى الربيعي لمجموعة من القيم (7.5) والمئين 75 لها (23) جد الربيع الأول لهذه المشاهدات؟

15- إذا كان مجموع مربعات (65) قيمة هي (4875) وانحرافها المعياري $\sqrt{11}$. احسب الوسط الحسابي لهذه القيم؟

16- للمفردات x1, x2,...,xn بين أن:

أ- $\sigma^2 = m_2\left(\bar{x}\right)$

ب- $L_3 = \frac{m_2\left(\bar{x}\right)}{\sigma^2}$

4

الوحدة الرابعة

الارتباط والانحدار

Correlation and Regression

الارتباط والانحدار

Correlation and Regression

المقدمة

الارتباط: هو قوة العلاقة بين متغيرين أو أكثر. ولكن سيقتصر تعاملنا في هذا الكتاب على متغيرين فقط.

جداول الانتشار وعلاقتها بالارتباط Scatter Tables

الانتشار هو التمثيل البياني للعلاقة بين المتغيرين ويكون ذلك برصد نقاط المتغيرين على المحورين الأفقي والعمودي.

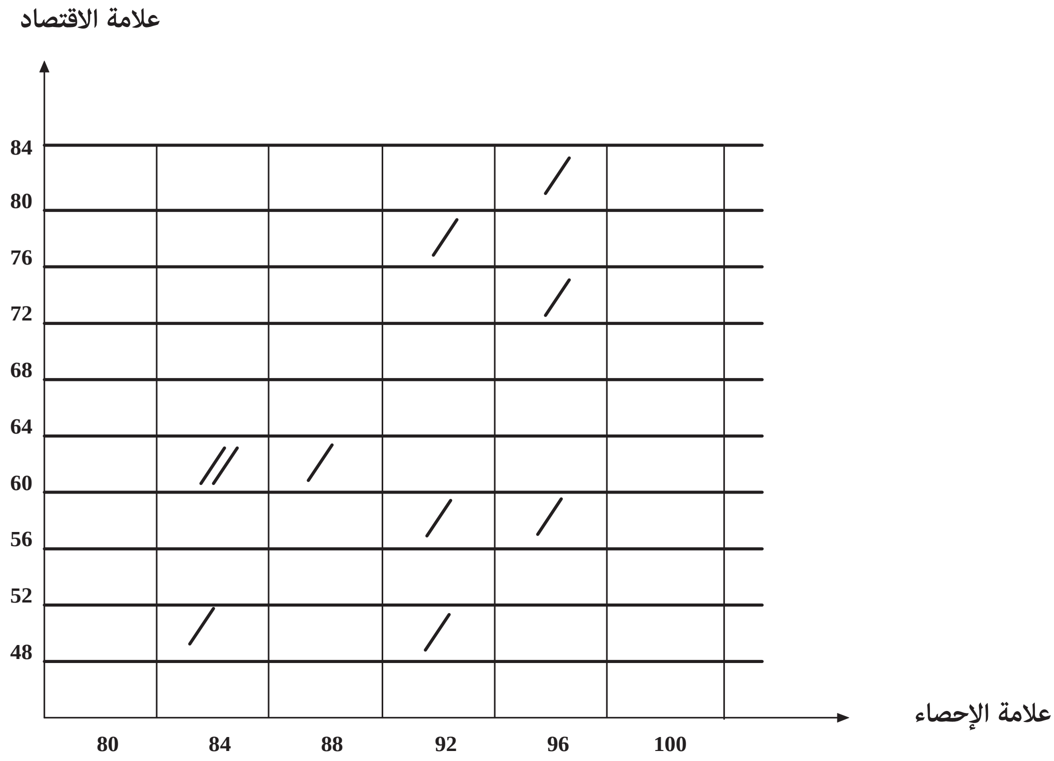
ومن خلال لوحة الانتشار يمكن تحديد مدى قوة العلاقة بين المتغيرين سواء كانت هذه العلاقة طردية (إيجابية) أو عكسية (سلبية).

مثال:

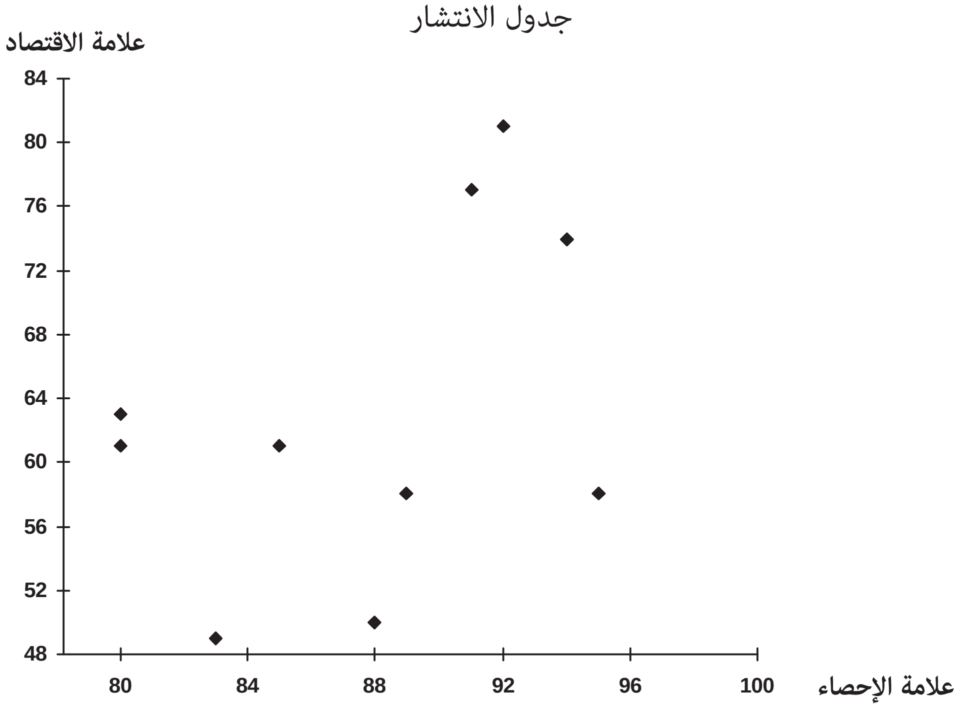
في دراسة أجريت لقياس مدى العلاقة بين التحصيل في مادة الإحصاء ومادة الاقتصاد، أجري امتحان في المادتين لعشرة طلاب وكانت النتائج كما يلي:

الإحصاء	الاقتصاد
80	61
95	58
94	74
89	58
92	81
83	49
91	77
88	50
80	63
85	61

كون جدول الانتشار ولوحة الانتشار وبـين فـيما إذا كانـت العلاقـة قويـة أم ضـعيفة وإذا كانـت سالبة أم موجبة.



جدول الانتشار



نرى من خلال لوحة الانتشار وجدول الانتشار أن الارتباط موجب ولكنه ضعيف نوعا ما ولكن إذا أردنا معرفة قوة العلاقة الحقيقية بين المتغيرين فإن أفضل مقياس هو معامل الارتباط والذي سنتحدث عنه في البند اللاحق.

معامل الارتباط Correlation Coefficient

وهو المقياس الرقمي لقوة الارتباط بين متغيرين.

خصائص معامل الارتباط Properties of correlation coefficient

1- تتراوح قيمة معامل الارتباط (r) بين 1- و 1.
أي $-1 \leq r \leq 1$

2- إذا كانت r بين 0 و (1) فإن العلاقة بين المتغيرين تكون علاقة موجبة أو طردية أما إذا كانت r بين 1- , 0 تكون العلاقة عكسية أو سالبة.

3- إذا كانت r =1 فإن العلاقة تكون موجبة تامة. وإذا كانت r = -1 فإن العلاقة سالبة تامة.

4- إذا كانت r = 0 فإنه لا يوجد علاقة بين المتغيرين.

5- تزداد قوة العلاقة كلما اقتربنا من (1) و (-1) وتقل كلما اقتربنا من 0.

وهناك عدة أنواع من معاملات الارتباط ولكن ستقتصر دراستنا في هذا الكتاب على معاملي ارتباط بيرسون وسبيرمان.

1- معامل ارتباط بيرسون Persons Correlation Coefficient: ويسمى معامل ارتباط العزوم. يعد من أفضل معاملات الارتباط وأكثرها شيوعاً.

(x1 , y1) , (x2, y2), …, (xn, yn)

فإن معامل ارتباط بيرسون هو:

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

حيث:

$\bar{x}$ = الوسط الحسابي لمشاهدات المتغير x.

$\bar{y}$ = الوسط الحسابي لمشاهدات المتغير y.

σ_x = الانحراف المعياري للمتغير x.

σ_y = الانحراف المعياري للمتغير y.

مثال:

أوجد معامل ارتباط بيرسون بين قيم x,y من البيانات التالية وبين فيما إذا كانت العلاقة طردية أم عكسية، ضعيفة أم قوية.

X	150	162	180	160	170	180
Y	200	250	300	200	240	280

الحل:

xi	yi	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x}).(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
150	200	-17	-45	765	289	2025
162	250	-5	5	-25	25	25
180	300	13	55	715	169	3025
160	200	-7	-45	315	49	2025
170	240	3	-5	-15	9	25
180	280	13	35	455	169	1225
1002	1470			2210	710	8350

نجد أولاً : $\bar{x} = 167 \quad , \quad \bar{y} = 245$

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} = \sqrt{\frac{710}{6}} = 10.88$$

$$\sigma_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}} = \sqrt{\frac{8350}{6}} = 37.31$$

$$r = \sum_{i=1}^{n}\frac{(x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x\sigma_y} = \frac{2210}{6*10.88*37.31} = 0.907$$

وهذا يدل على أن العلاقة طردية قوية.

وهناك أشكال أخرى لمعادلة إيجاد معامل ارتباط بيرسون منها.

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

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

مثال:

احسب معامل ارتباط بيرسون بين علامات عشرة طلاب في مساقي الرياضيات والإحصاء من الجدول التالي:

x	80	60	55	40	75	85	70	60	80	90
y	75	65	60	50	70	90	70	55	80	85

الحل:

ستستخدم القانون

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

الرياضيات (x)	**الإحصاء** (y)	*xy*	*x2*	*y2*
80	75	6000	6400	5625
60	65	3900	3600	4225
55	60	3300	3025	3600
40	50	2000	1600	2500
75	70	5250	5625	4900
85	90	7650	7225	8100
70	70	4900	4900	4900
60	55	3300	3600	3025
80	80	6400	6400	6400
90	85	7650	8100	7225
695	700	50350	50475	50500

$$\bar{x} = 69.5$$

$$\bar{y} = 70$$

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

$$= \frac{50350 - 10 * 69.5 * 70}{\sqrt{50475 - 10(69.5)^2}\ \sqrt{50500 - 10(70)^2}}$$

$$= \frac{1700}{\sqrt{2172.5}\ \sqrt{1500}} = \frac{1700}{1805.2} = 0.942$$

الارتباط (طردي قوي)

2- معامل ارتباط سبيرمان Sparman Correlation Coefficient:

ويسمى معامل ارتباط الرتب

يستخدم إذا كان هناك صعوبة في استخدام معامل ارتباط بيرسون أو لم يتوفر لدينا القيم الحقيقية للمشاهدات ولكن توفرت رتبها فقط ويكون:

$$\boldsymbol{r} = 1 - \frac{6\sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \boldsymbol{d}_{\boldsymbol{i}}^2}{\boldsymbol{n}(\boldsymbol{n}^2 - 1)}$$

حيث di: فروق بين رتب المتغيرين x , y.

أي di = Ox - Oy

حيث Ox : رتبة x

Oy : رتبة y

وبشكل مختصر

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