الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التكوين والتعليم المهنيين
Ministère de la Formation et de l'Enseignement Professionnels

Centre National de l'Enseignement Professionnel à Distance (CNEPD)

المركز الوطني للتعليم المهني عن بعد (م.و.ت.م.ب)

درس الإحصاء

السلسلة رقم05

التعديل و الإرتباط

خطة الدرس :

I- التوزيعات الإحصائية المزدوجة

II- التعديل (عموميات)

1- التعديل البياني
2- التعديل الآلي
3- التعديل التحليلي

III- الإرتباط

1- تصنيف السلاسل الإحصائية بمتغيرين
2- تعريف الإرتباط
3- مستقيمات الإنحدار
4- معامل الإرتباط

WWW.DZPEN.COM | قلم الجزائر

IV- تحديد مستقيمات الانحدار

1- مستقيم الإنحدار للعينات على محور السينات : $D_y(x)$

2 - مستقيم الإنحدار للسينات على محور العينات: $D_x(y)$

V- معامل الارتباط الخطي

VI- صيغ أخرى لمعاملات توجيه مستقيمات الإنحدار

VII- تفسير قيمة معامل الإرتباط ووضعية مستقيمات الإنحدار

1- ملاحظة

2- نتائج أولية

3- حوصلة النتائج

WWW.DZPEN.COM | قلم الجزائر

I– التوزيعات الإحصائية المزدوجة :

يمكننا أن نتطرق الآن إلى الملاحظات التي تنصب في نفس الوقت على متغيرين مهما كانت طبيعتهما كمية أو نوعية. و هذه الملاحظات تعرض في جداول إحصائية "مزدوجة" أي مدخل لكل متغير.

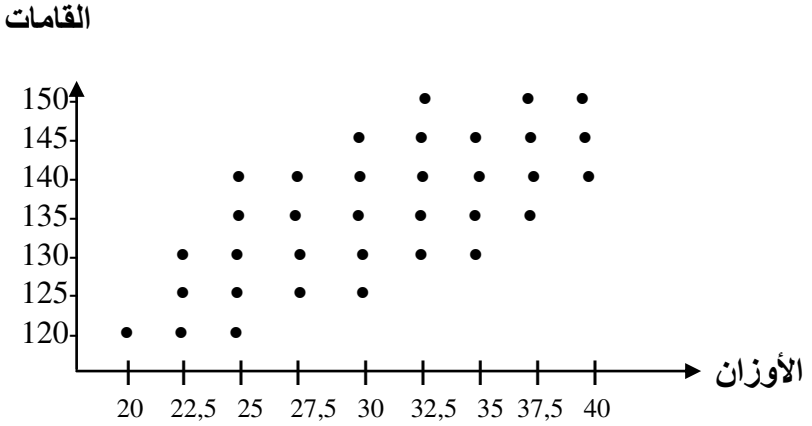
مثال :

ليكن التوزيع التالي حسب القامة و الوزن الذي يخص 300 طفل، عمر كل واحد منهم 9 سنوات.

Y_i / x_i الأوزان (كغ)	القامات (سم)							كل القامات
	120	125	130	135	140	145	150	
20	4	--	--	--	--	--	--	4
22,5	3	7	2	--	--	--	--	12
25	2	17	20	7	1	--	--	47
27,5	--	8	30	28	6	--	--	72
30	--	1	17	36	19	4	--	77
32,5	--	--	4	16	22	3	2	47
35	--	--	1	5	10	8	--	24
37,5	--	--	--	1	4	5	2	12
40	--	--	--	--	2	1	2	5
كل الأوزان (كغ)	**9**	**33**	**74**	**93**	**64**	**21**	**6**	**300**

التمثيل البياني :

لنرفق كل ملاحظة بنقطة M_i إحداثياتها (x_i , y_i) :



II – التعديل : عموميات:

إن التمثيل البياني لسلسلة إحصائية بمتغيرين تبين:

– أنه لكل قيمة حقيقية x_i من قيم المتغير يرفق تكرار n_i (أو تكرار نسبي f_i) و هذه الإحداثيات تعين نقطة p_i بحيث يتم عرض الظاهرة الملاحظة على المستوى بمجموعة من النقاط تشكل سحابة نقطية. و هذه النقاط يمكن ربطها بخط منكسر.

– و أن التعديل يعني إنشاء منحنى بحيث يكون بين كل قيمتين حقيقيتين للمتغير، مجموعة غير منتهية من النقاط (استكمال داخلي). و حيث يكون ، قبل و بعد القيم الحقيقية المتطرقة للمتغير، مجموعتان غير منتهيتين من النقاط (استكمال خارجي).

و أن يكون هذا المنحنى أقرب ما يمكن من النقاط الحقيقية p_i و ذلك للإلتزام بشروط الإنتظام.

WWW.DZPEN.COM | قلم الجزائر

وبذلك، يكون لكل قيمة حقيقية x_i للمتغير نقطة مطابقة p_i من المنحنى و التي من المفروض أن يكون ترتيبها y'_i أكثر طبيعيا من الترتيب y_i الموافق للنقطة p_i التي لها نفس الفاصلة.

و هكذا فإن التعديل يؤدي إلى رسم منحنى يمكن تحديده بواسطة:

- تعديل بياني؛
- تعديل آلي ؛
- تعديل تحليلي.

و لهذا سندرس فيما يلي مختلف أنواع التعديلات:

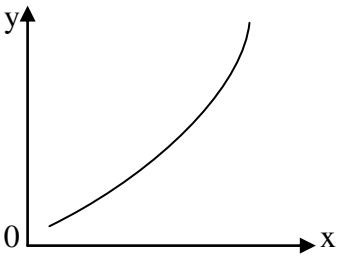
1- <u>التعديل البياني</u>:

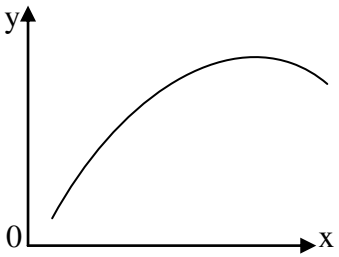
إنطلاقا من البيان الذي يمثل مختلف الملاحظات (سحابة من النقاط)، نرسم بيد مرفوعة منحنى منتظما بقدر الإمكان يمر عبر سحابة النقاط بأحسن كيفية.

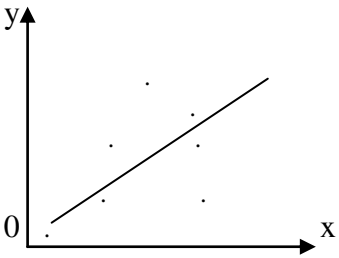
و هذا المنحنى الذي يتم تعديله بهذه الطريقة قد يكون مستقيما. وعندئذ نقول عنه أنه تعديل خطي.

عموما بهذه الطريقة تتأثر كثيرا بالشخصية. لكن التجربة تبين انه عندما يقوم عدة إحصائيين بتعديل منحنى، على أساس نفس المعطيات، فإنهم يحصلون على نتائج متقاربة جدا.

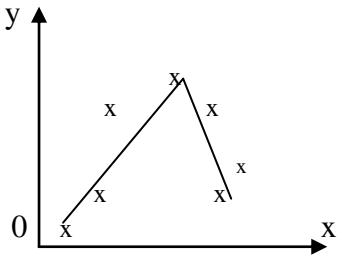
أمثلة :







تعديل سيء



تعديل جيد

2- التعديل الآلي:

في هذه الحالة ، تكون كل قيمة معدلة $\bar{y}$ (حسب القيم y_i) دالة للقيم المشاهدة بجوار القيمة x_i كما أن القيم $\bar{y}_i$ هي المتوسطات الحسابية للقيم y_i المعنية.

و لهذا يتم تعويض كل نقطة بنقطة أخرى p_i فاصلتها x_i و ترتيبها y_i وهكذا يمكن إجراء التعديل الآلي بطريقتين هما:

- طريقة المتوسطات المجدولة؛
- طريقة المتوسطات المتحركة.

WWW.DZPEN.COM | قلم الجزائر

1.2- طريقة المتوسطات المجدولة:

في البداية ، نقسم كل النقاط n_i الموافقة لأزواج المعطيات (x_i ,y_i) إلى مجموعات جزئية منفصلة بحيث تتكون كل واحدة منها من نفس العدد من النقاط على أن يكون هذا العدد فرديا.

ثم نميز مجموعة جزئية بواسطة قيمتين هما:

- قيمة x_i و هي تساوي الوسيط لقيم المتغير؛
- قيمة معدلة y_i ، و هي تساوي المتوسطة الحسابي للقيم y_i.

مثل :

لنعتبر السلسلة الإحصائية التالية :

x_i	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150
y_i	6	7	10	15	16	20	29	30	32	36	37	44	45	49	50

P□3 **P□8**

- يمكن تمثيل المجموعة الجزئية الأولى بالزوج:

$$x_3 = 30 \quad \text{و} \quad y'_3 = \frac{6+7+10+15+16}{5} = 10,8$$

و بالنقطة p'_3

- و يكون تمثيل المجموعة الجزئية الثانية بالزوج:

$$x_8 = 80 \quad \text{و} \quad y'_8 = \frac{20+29+30+32+36}{5} = 29,4$$

و بالنقطة p'_8.

- بينما نمثل المجموعة الجزئية الثالثة بالزوج:

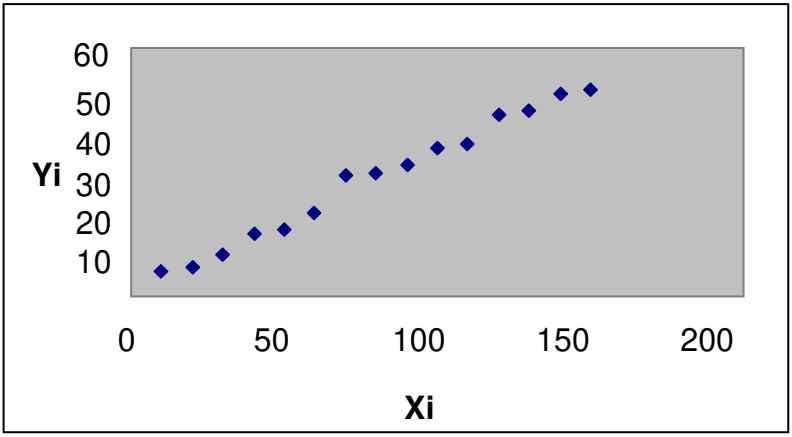
$$x_{13} = 130 \quad \text{و} \quad y'_{13} = \frac{37+44+45+49+50}{5} = 45$$

WWW.DZPEN.COM | قلم الجزائر

و بالنقطة p'_{13} .

و في النهاية نحصل على المنحنى المعدل ، و هو المنحنى الذي يربط النقاط

$P`_3$ , $P`_8$, $P`_{13}$.



2.2- طريقة المتوسطات المتحركة:

بداية نقسم مجموعة النقاط n_i الموافقة لإحداثيات المعطيات كما فعلنا في المتوسطات المجدولة ، إلى مجموعات جزئية على أن تكون هذه المجموعات الجزئية غير منفصلة.

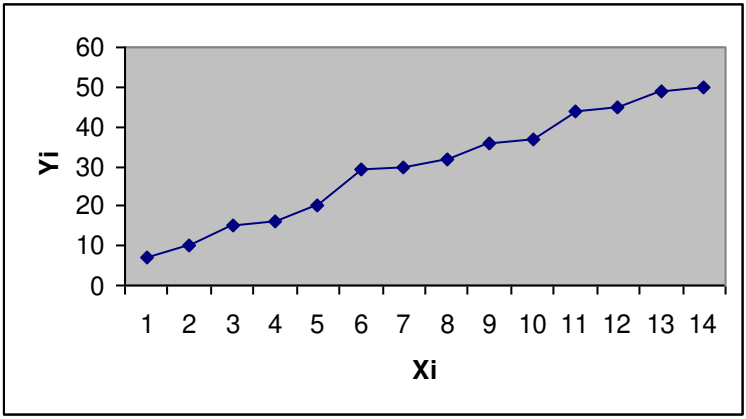
و عند تجميع "k" زوج من المعطيات (من الأفضل أن يكون "k" فرديا لتسهيل عملية حساب الوسيط)، فإن أي مجموعة من المجموعات الجزئية المتحصل عليها بهذه الكيفية تتكون من (k-1) من إحداثيات المعطيات الأخيرة الخاصة بالمجموعة الجزئية السابقة و من زوج المعطيات الذي يلي مباشرة المجموعة الجزئية السابقة.

ثم نميز كل مجموعة جزئية كما فعلنا سابقا: بقيمة x_i و هي تساوي الوسيط للقيم الخاصة بالمجموعة الجزئية.

بقيمة معدلة y_i، و هي تساوي المتوسط الحسابي للقيم y_i المقابلة لها.

لنسترجع المثال السابق و لنحسب متوسطات متحركة خمسة بخمسة (k=5).

x_i	y_i	متوسطات مجدولة y_i (لكل 5 مفردات)	متوسطات متحركة y_i (بـ 5 مفردات)
10	6		
20	7		
30	10	10,8	(6+7+10+15+16)1/5 = 10,8
40	15		(7+10+ 15+ 16+20)1/5 = 13,6
50	16		(10+15+16+20+29)1/5 = 18
60	20		(15+16+20+29+30)1/5 = 22
70	29		(16+20+29+30+32)1/5=25,4
80	30	29,4	(20+29+30+32+36) =29,4
90	32		(29+30+32+36+37)= 32,4
100	36		(30+32+36+37+44) =35,8
110	37		(32+36+37+44+45) = 38,8
120	44		(36+37+44+45+49) = 42,2
130	45	45	(37+44+45+49+50)1/5 = 45
140	49		
150	50		



WWW.DZPEN.COM | قلم الجزائر

3- التعديل التحليلي:

تهدف هذه الطريقة إلى إنشاء منحنى التقدير (مستقيم) إعتمادا على معاينة سحابة النقاط التي تمثل الملاحظات (x_i, y_i).

في هذه الحالة نقول عن مسألة التعديل أنها خطية، و يمكن تحديد مستقيم التقدير بواسطة إحدى الطرق العامة المذكورة أعلاه، أو بواسطة طريقة المربعات الصغرى (و هي الطريقة الأكثر إستعمالا).

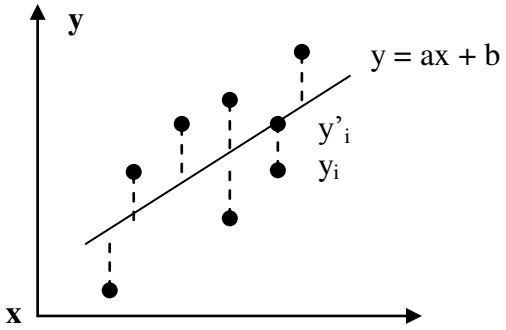
1.3- طريقة المربعات الصغرى:

نفترض أن منحنى تقدير y بدلالة x يمكن تمثيله بمستقيم، و حيث أن معادلة المستقيم المعدل هي من الشكل: $y = ax + b$.

علينا أن نحدد قيمتين a و b بحيث يكون مجموع مربعات الفروق بين القيم الملاحظة y_i و القيم المعدلة y'_i صغير.

أي أن يكون : $\sum_{i=1}^{n}[y_i - y'_i]^2$ صغير.

و من هذا الشريط يأتي إسم طريقة المربعات الصغرى.



2.3- معادلة مستقيم تقدير y بدلالة x:

إن معادلة هذا المستقيم هي من الشكل y = ax + b ، على أن يتم تحديد قيمتي العددين a و b بحيث أن:

$$A = \sum_{i=1}^{n} [y_i - y'_i]^2$$

في العبارة : $A = \sum_{i=1}^{n} [y_i - y'_i]^2$ نجد أن x_i و y_i هما عددان معلومان (قيم مشاهدة) ، يجب علينا حساب قيمتي a و b.

$$A = \sum_{i=1}^{n} [(y_i - ax_i - b)]^2 =$$

$$\sum_{i=1}^{n} \left({y_i}^2 - 2\ ax_i y_i - 2by_i + a^2 {x_i}^2 + 2abx_i + b^2 \right)$$

رياضيا نبرهن أنه يكون لدالة نهاية صغرى بالنسبة للثنائية (a,b) ، عندما ينعدم المشتق الجزئي للعبارة A بالنسبة للعنصر a ، و كذلك مشتقتها بالنسبة للعنصر b.

هذان المشتقان هما كالتالي :

$$A`a = \sum_{i=1}^{n} -2x_i y_i + \sum_{i=1}^{n} 2a{x_i}^2 + \sum_{i=1}^{n} 2bx_i$$

$$A`a = -2\sum_{i=1}^{n} \left(x_i y_i - a{x_i}^2 - bx_i \right) = -2\sum_{i=1}^{n} x_i (y_i - ax_i - b)$$

$$A`b = \sum_{i=1}^{n} -2y_i + \sum_{i=1}^{n} 2ax_i + \sum_{i=1}^{n} 2b = -2\sum_{i=1}^{n} (y_i - ax_i - b)$$

ينعدم هذان المشتقان عندما يكون:

$$A`a = -2\sum_{i=1}^{n} x_i (y_i - ax_i - b) = 0 \ ; \ \sum_{i=1}^{n} x_i (y_i - ax_i - b) = 0 \qquad \textbf{(1)}$$

$$A`b = -2\sum_{i=1}^{n} (y_i - ax_i - b) = 0 \ ; \ \sum_{i=1}^{n} (y_i - ax_i - b) = 0 \qquad \textbf{(2)}$$

WWW.DZPEN.COM | قلم الجزائر

نلاحظ أن المعادلة II لا تقدم شيئا جديدا، لأنها تعبر على ماتم البرهنة عليه في المعادلة I.

أ- **تحديد معاملي المستقيم : (الحالة العامة):**

لنحسب قيمة b من المعادلة I أو II كمايلي:

$\bar{y} = \overline{ax} + b \Rightarrow b = \bar{y} - \overline{ax}$

و بتعويض قيمة b في المعادلة I :

$$\sum_{i=1}^{n} x_i(y_i - ax_i - b) = 0; \sum_{i=1}^{n} x_i(y_i - ax_i - \bar{y} + a\bar{x}) = 0$$

$$\sum_{i=1}^{n} x_i\left[y_i - \bar{y} - a(x_i - \bar{x})\right] = 0$$

$$\sum_{i=1}^{n} x_i y_i - \overline{n}\,\mathrm{x}\,\bar{y} - a\sum_{i=1}^{n} x_i^{\,2} + na\bar{x}^{2} = 0$$

$$\sum_{i=1}^{n} x_i y_i - \overline{n} \times \bar{y} - a(\sum_{i=1}^{n} x_i^{\,2} + n\bar{x}^{2}) = 0$$

و منه :

$$a = \frac{\sum_{i=1}^{n} x_i y_i - \overline{n}\,\bar{y}}{\sum_{i=1}^{n} x_i^{\,2} - n\bar{x}^{2}}$$

هذه الصيغة تسمى صيغة المنشورة

بعد تحديد قيمة "a" ، نعوض a و $\bar{y}$ و $\bar{x}$ بقيمتها في المعادلة :

$\bar{y} = a\bar{x} + b$، و ذلك للحصول على قيمة "b".

ب- تحديد قيمة المعامل a بواسطة المتغيرات المركزة:

نحصل على: $a = \dfrac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^{\,2} - n\bar{x}^{2}}$ من الصيغة:

$$a = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^{n} x_i^{\,2} - \bar{x}^{2}} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y} - \bar{x}\bar{y} + \bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^{n} x_i^{\,2} - 2\bar{x}^{2} + \bar{x}^{2}}$$

$$a = \frac{\frac{1}{n}\sum_{i=1}^{n} (x_i y_i - \bar{x}\bar{y} - \bar{x}\bar{y} + \bar{x}\bar{y})}{\frac{1}{n}\sum_{i=1}^{n} (x_i^{\,2} - 2x_i\bar{x} + \bar{x}^{2})} = \frac{\sum_{i=1}^{n} (x_i y_i - \bar{x}\bar{y} - \bar{x}\bar{y} + \bar{x}\bar{y})}{\sum_{i=1}^{n} x_i^{\,2} - 2x_i\bar{x} + \bar{x}^{2}}$$

$$a = \frac{\sum_{i=1}^{n} \left[(x_i - \bar{x})(y_i - \bar{y}) \right]}{\sum_{i=1}^{n} (x_i - \bar{x})^{2}}$$

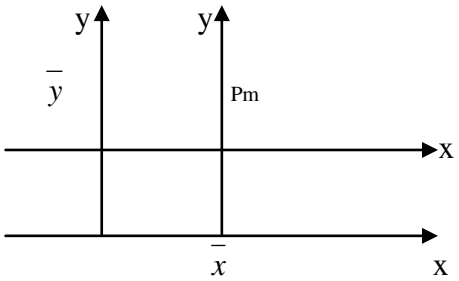
ج- حساب قيمة a مباشرة :

لنغير مبدأ الاحداثيات بحيث يكون للعنصر:

x_i يوافق $x_i - \bar{x} = x_i$

y_i يوافق $y_i - \bar{y} = y_i$

y'_i يوافق $y'_i - \bar{y} = y'_i$



بهذا يكون مبدأ الاحداثيات الجديد هو النقطة المتوسطة $n_i(\bar{x}\,،\,\bar{y})$.

و تصبح معادلة المستقيم المعدل من الشكل : $y` = ax$.

و بالتعويض في الصيغة السابقة نجد أن:

$$a = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} {x_i}^2} \qquad \Rightarrow y' = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} {x_i}^2}$$

مثال : لتكن السلسلة الإحصائية السابقة:

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
10	6	-70	-22,4	4900	+ 1568
20	7	-60	-21,4	3600	+ 1284
30	10	-50	-18,4	2500	+ 920
40	15	-40	-13,4	1600	+ 536
50	16	-30	-12,4	900	+ 372
60	20	-20	- 8,4	400	+ 168
70	29	-10	0,6	100	- 6
80	30	0	1,6	0	0
90	32	10	3,6	100	+ 36
100	36	20	7,6	400	+ 152
110	37	30	8,6	900	+ 258
120	44	40	15,6	1600	+ 624
130	45	50	16,6	2500	+ 830
140	49	60	20,6	3600	+ 1236
150	50	70	21,6	4900	+ 1512
1200	**426**			**28000**	**+9490**

WWW.DZPEN.COM | قلم الجزائر

$$\bar{x} = \frac{1200}{15} = 80 \qquad \bar{y} = \frac{426}{15} = 28,5$$

$$a = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{9490}{28000} = 0,338$$

$b = \bar{y} - a\bar{x}$; $b = 28,4 - 0,338 \times 80 = 1,36$

y = 0,338 x + 1,36

III- الإرتباط :

في الفقرات السابقة شرحنا كيف يمكن تعديل منحنى التقدير الذي يمثل التغير المتوسط للتغير المفسر y بدلالة المتعبير المفسر x .

لكن يتعين الآن أن نوضح كيف تتوزع الملاحظات بالنسبة لمنحنى التقدير و على وجه التدقيق سنهتم بتقييم كيفية إنتشار سحابة النقاط التي تمثل الملاحظات ، بالنسبة إلى منحنى التقدير . وبتعبير آخر قياس الإرتباط، أي قياس شدة الترابط الإحصائي بين المتغيرين.

1- تصنيف سلاسل إحصائية بمتغيرين:

يمكن تصنيف السلاسل الإحصائية بمتغيرين إلى ثلاثة أصناف:

1.1- السلاسل التي تكون متغيراتها مترابطة تابعيا :

مثال:

السلاسل التي تخص أقطار الدوائر و محيطاتها ، حيث يتم التعبير عن إرتباط المتغيرين بعلاقة رياضية ، فالقطر يكون بدلالة المحيط و العكس صحيح أي أنها علاقة تابعية.

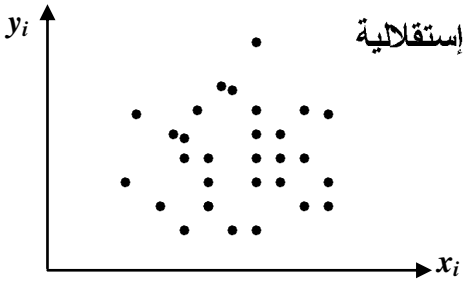
و لتمثيل هذه العلاقة بيانيا نسجل قيم الأقطار x على محور الفواصل و قيم المحيطات y على محور التراتيب ،عندئذ نحصل على سلسلة إحصائية يمثلها مستقيم.

2.1- السلاسل التي تكون متغيراتها مستقلة:

مثال:

لتكن السلسلة التي تخص قامة العمال و أجورهم في إحدى المؤسسات.و بديهي أنه لا يوجد بين هذين المتغيرين أي إرتباط بل إنهما مستقلان.

و عند تمثيلها بيانيا، تنتشر النقاط التي تجسم السحابة بصورة عشوائية على المستوي.



3.1- السلاسل التي تكون متغيراتها مترابطة جزئيا :

مثال :

- قامة و وزن التلاميذ في أحد الأقسام؛
- إنتاجية العمال و أجورهم في إحدى المؤسسات؛
- عدد حوادث المرور و نسبة تناول الكحول عند السائقين.

في كل حالة من هذه الحالات يوجد ترابط إحصائي.

و عند تمثيلها بيانيا لا تكون النقاط المجسمة للسحابة واقعة بجوار قريب من المنحنى، و لا تكون منتشرة على المستوي، و إنما تتوزع على جزء من هذا المستوي.

WWW.DZPEN.COM | قلم الجزائر

و حسب إتجاه سحابة النقاط، يمكن أن نعرف إن كانت قيم المتغيرين تتغير بنفس الإتجاه، أو تتغير بإتجاهين متعاكسين.

و بالتالي نقول: هناك إرتباط.



2- تعريف الإرتباط:

إن الإرتباط هو تعبير عن شدة الترابط التابعي بين متغيرين إحصائيين.

ملاحظات:

- يكون الترابط التابعي أكبر شدة بقدر ما يكون شكل السحابة أكثر ضيقا، و بالتالي شبيها بمنحنى.
- إن هذه السحابة قد تكون على أشكال متعددة، سواء كانت أشكال هندسية مألوفة أم لا.
- سنقتصر على دراسة السحابة التي لها شكل القطع الناقص الذي قد يكون متزايد أو متناقص.

و من ثمة، يمكن اللجوء إلى تعديل خطي يميز أبسط الإرتباطات، و هو الإرتباط الخطي.

3- مستقيمات الإنحدار:

لإجراء التعديل يبحث الإحصائي، بواسطة طريقة المربعات الصغرى عن معادلة المستقيم بحيث يكون للقيم $y`_i$ المعدلة على طول هذا المستقيم إنطلاقا من القيم x_i، إقتراب جيد من القيم الحقيقية y_i.

WWW.DZPEN.COM | قلم الجزائر

و هذا المستقيم نسميه: مستقيم الإنحدار للمتغير y بالنسبة للمتغير x أو مستقيم التقدير للعينات بدلالة x، و نرمز له بالرمز $D_y(x)$.

و فيما يخص السلاسل بمتغيرين ، قد يكون للتعديل العكسي مدلول يعادل مدلول التعديل المباشر ، و لهذا يمكننا إيجاد مستقيم أخر، هو $D_x(y)$ أو مستقيم الإنحدار للمتغير x بالنسبة إلى المتغير y.

4- معامل الإرتباط :

إن الإرتباط بين قيم متغيرين أكثر شدة بقدر إقتراب مستقيمي الإنحدار من بعضها البعض، و بالتالي يمكن قياس شدة الإرتباط بواسطة عدد يتوقف على و ضعيتي المستقيمين بالنسبة لبعضها البعض. و في هذا العدد نسميه: معامل الإرتباط الخطي.

IV- تحديد مستقيمات الإنحدار:

1- مستقيم إنحدار العينات على محور السينات : $D_y(x)$.

لنرمز إلى القيم $x_i - \bar{x}$ بالرمز x_i .
ولنرمز إلى القيم $y_i - \bar{y}$ بالرمز y_i .

و حيث سبق أن رأينا أن مستقيم الإنحدار يمر بالنقطة المتوسطة P_m التي لها الإحداثيات $\bar{x}$ و $\bar{y}$ ، و أن صيغة معامل توجيهه هي:

$$a_y(x) = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

$$y - \bar{y} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}(x - \bar{x})$$

WWW.DZPEN.COM | قلم الجزائر

أو:

و بالتالي نكتب:

$$y = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} x$$

2- مستقيم إنحدار السينات على محور العينات: $D_x(y)$

في هذه الحالة، يتعلق الأمر بتعويض القيمة المعدلة x_i` محل القيمة x_i الموافقة لمتغير محدد y_i.

$$a_x(y) = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^2}$$

$$x = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^2} y \quad \Rightarrow y = \frac{1}{a_x(y)} x$$

$$x - \bar{x} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^2} (y - \bar{y}) \quad \Rightarrow \quad y - \bar{y} = \frac{1}{a_x(y)} (x - \bar{x})$$

تطبيق:

لتكن سلسلة إحصائية تتكون من قامات و أوزان تلاميذ قسم ما.

القامات x_i	الأوزان y_i	$X_i = x_i - \bar{x}$	$Y_i = y_i - \bar{y}$	$X_i Y_i$	X_i^2	Y_i^2
154	54	- 10	- 12	120	100	144
155	58	- 9	- 8	72	81	64
156	61	- 8	- 5	40	64	25
160	63	- 4	- 3	12	16	9
162	65	- 2	- 1	2	4	1
162	61	- 2	- 5	10	4	25
167	68	+ 3	+ 2	6	9	4
171	68	+ 7	+ 2	14	49	4
173	76	+ 9	+ 10	90	81	100
180	86	+ 16	+ 20	320	256	400
1640	**660**			**686**	**664**	**776**

$$\bar{x} = \frac{1640}{10} = 164 \quad ; \quad \bar{y} = \frac{660}{10} = 66$$

$$a_y(x) = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} = \frac{686}{664} = 1{,}033$$

$$a_x(y) = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^2} = \frac{686}{776} = 0{,}884$$

$$\frac{1}{a_x(y)} = \frac{1}{0{,}884} = \frac{776}{686} = 1{,}131$$

- معادلة الإنحدار للعناصر y بدلالة العناصر x :

$$y - 66 = 1{,}033\,(x - 164)$$

أي:

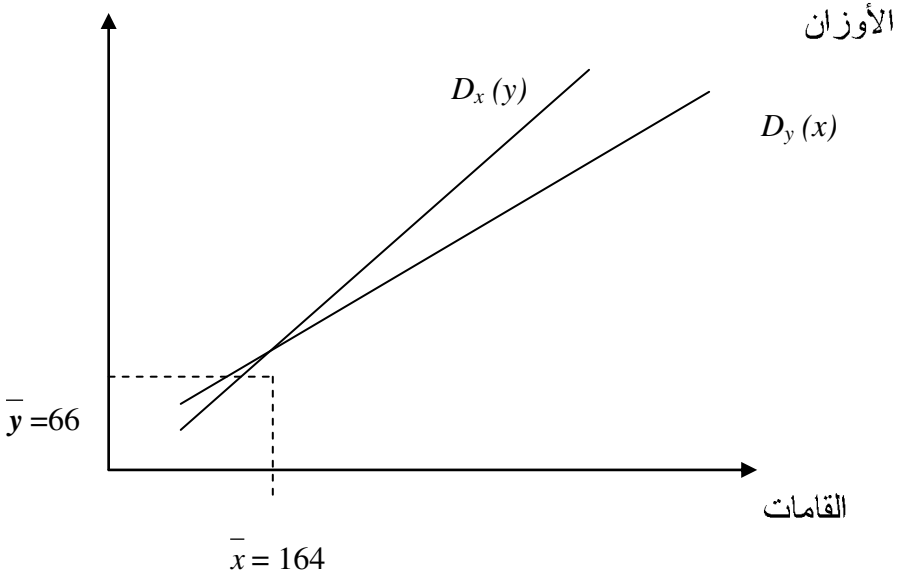
$$y = 1{,}033\,x - 103{,}412$$

- معادلة الإنحدار للعناصر x بدلالة العناصر y :

$$y - 66 = 1{,}131\,(x - 164)$$

أي:

$$y = 1{,}131\,x - 119{,}484$$



V- معامل الإرتباط الخطي:

إن معامل الإرتباط الخطي هو مقياس لشدة الإرتباط بين متغيرين، و صيغته هي:

$$\lambda = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i}{\delta_x \times \delta_y}$$

$$\lambda = \frac{\sum_{i=1}^{n} x_i y_i}{n\delta_x . \delta_y}$$

حيث :

σ_x : الإنحراف المعياري للمتغير x.

σ_y : الإنحراف المعياري للمتغير y.

ملاحظة : نعلم أن σ_x و σ_y هما مقداران موجبان، و كذلك المقدار n ، و بالتالي فإن مقام λ هو حتما مقدار موجب. و ، لهذا فإن إشارته تتوقف على إشارة المجموع : $\sum_{i=1}^{n} x_i y_i$.

VI – صيغ أخرى لمعاملات توجيه مستقيمات الإنحدار:

لقد برهنا على أن: $\lambda = \frac{\sum_{i=1}^{n} x_i y_i}{n\delta_x . \delta_y}$

كما أننا نعلم أن: $a_y(x)\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_n^2} = \frac{\sum_{i=1}^{n} x_i y_i}{n\delta_x^2}$

وأن:

$$a_x(y) = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^2} = \frac{\sum_{i=1}^{n} x_i y_i}{\lambda \delta_y^2}$$

بالتالي :

$$a_x(y) = \lambda \frac{\delta x}{\delta y} \qquad a_y(x) = \lambda \frac{\delta_y}{\delta x}$$

و من هذه النتائج نستخلص صيغة جديدة لمعامل الإرتباط الخطي هي:

$$\lambda = \frac{\sum_{i=1}^{n} x_i y_i}{n \delta_x . \delta_y} \Rightarrow \lambda^2 = \frac{\sum_{i=1}^{n} [x_i y_i]^2}{n^2 \delta_x{}^2 \delta_y{}^2} = \frac{\sum_{i=1}^{n} x_i y_i}{n \delta_x^2} \times \frac{\sum_{i=1}^{n} x_i y_i}{n \delta_y^2}$$

و منه :

$$\lambda^2 = a_y(x) \times a_x(y)$$

و هذا يعني أن مربع عامل الإرتباط الخطي يساوي جداء معاملي التوجيه لمستقيمي الإنحدار.

لنسترجع المثال السابق و لنحسب قيمة λ :

$$\delta_x^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} = \frac{664}{10} = 6{,}64 \Rightarrow \delta x = 8{,}14$$

$$\delta_y^2 = \frac{\sum_{i=1}^{n} y_i^2}{n} = \frac{776}{10} = 77{,}6 \Rightarrow \delta y = 8{,}80$$

$$\lambda = \frac{\sum_{i=1}^{n} x_i y_i}{n\delta_x.\delta_y} = \frac{686}{10 \times 8{,}14 \times 8{,}80} = \boxed{0.95}$$

و بصيغة أخرى:

$$\lambda^2 = a_y(x) \times a_x(y) = 1{,}033 \times 0{,}884 = 0{,}9131$$

VII- تفسير قيمة معامل الإرتباط و وضعية مستقيمات الإنحدار:

1- ملاحظة:

يتحدد مستقيم الإنحدار $D_y(x)$ في المستوي (x,P_m,y) بحيث يكون مجموع مربعات "المتبقيات" d_i بين القيم المشاهدة y_i ، و القيم $a_y(x) \times x_i$ المعادلة على طول ذلك المستقيم الصغير.

و هذا المجموع يكتب:

$$\sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} (y_i - a_y(x)x_i)^2$$

$$\sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} y_i^2 - 2a_y(x)\sum_{i=1}^{n} x_i y_i + a_y^2(x)\sum_{i=1}^{n} x_i^2$$

لكننا نعلم أن :

$$a_y(x) = \frac{\sum_{i=1}^{n} x_i y_i}{n\delta_x^2} \quad \Rightarrow \quad \sum_{i=1}^{n} x_i y_i = n\delta_x^2 \times a_y(x)$$

بالتعويض يكون:

$$\sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} y_i^2 - 2\, a_y^2(x)\, n\, \delta_x^2 + a_y^2(x) \sum_{i=1}^{n} x_i^2$$

بما أن :

$$\sum_{i=1}^{n} x_i^2 = n\delta_x^2 \quad \text{و}$$

$$\sum_{i=1}^{n} y_i^2 = n\delta_y^2$$

يمكن أن نكتب:

$$\sum_{i=1}^{n} d_i^2 = n\delta_y^2 - 2a_y^2(x).n\delta_x^2 + a_y^2(x).n\delta_x^2$$

$$= n\ \delta_y^2 - a_y^2(x)\ .n\ \delta_x^2$$

$$= n\ (\ \delta_y^2 - a_y^2\ (x)\ \delta_x^2\ )$$

بما أن :

$$a_y(x) = \lambda \frac{\delta_y}{\delta_x} \Rightarrow \sum_{i=1}^{n} d_i^{\,2} = n(\delta_y^2 - \lambda^2\delta_y^2) = n\delta_y^2(1-\lambda^2)$$

و منه :

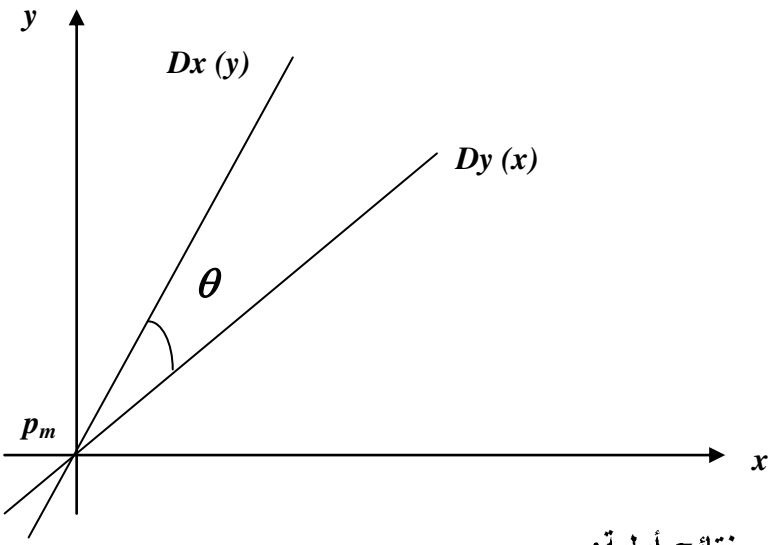
$$\frac{1}{n} \sum_{i=1}^{n} d_i^{\,2} = \delta_y^2(\ 1-\lambda^2\ ) = S_y^2$$

و هذه العبارة S_y^2 هي تباين البواقي d_i . الجذر التربيعي الموجب لهذا التباين هو S_y و نسميه "خطأ لقيمة y ".

و باستخدام نفس الإستبدلال يمكن كتابة "خطأ التقدير لقيمة x " كالتالي :

$$S_x^2 = \delta_y^2\ (1-\lambda^2).$$

و نلاحظ أن رسم المستقيمين $D_y(x)$ و $D_x(y)$ في نفس المعلم، يشكل زاوية نسميها θ "زاوية الإنحدار ".



2 – نتائج أولية:

1.2– مجال تغير λ:

$S^2x = \delta_x^2 (1- \lambda^2)$ و $Sy^2 = \sigma y^2 (1 - \lambda^2)$.

لدينا :

في العبارتين السابقتين تمثل $s_x^2, \delta_x^2, s_y^2, \delta_y^2$ التباينات و بالتالي فهي إما موجبة

أو معدومة. وهذا يستلزم $(1-\lambda^2)$ هو الآخر إما موجب أو معدوم. و بالتالي

$$-1 \leq \lambda \leq +1$$

نستنتج أن:

الحالة التي يكون فيها n = ± 1 : إنحدار أدنى و إرتباط أعظم:

عندما يأخذ معامل الإرتباط الخطي λ القيمة (+1)، أو القيمة (−1) يكون مستقيما الإنحدار $D_x(y)$، $D_y(x)$ منطبقان على بعضهما البعض و يكون بين المتغيرين x و y علاقة تابعية.

و عندما يكون مستقيما الإنحدار منطبقين تكون قيمة زاوية الإنحدار θ صفرا.

3.2- الحالة التي تكون فيها λ = 0 : إنحدار أعظم و إرتباط أدنى:

في هذه الحالة يكون مستقيما الإنحدار منطبقين على محوري الإحداثيات اللذين يمران بالنقطة المتوسطة P_m. و بالتالي تكون هناك إستقلالية بين قيم المقدارين.

3 - حوصلة للنتائج :

λ	-1 ← 0 ← +1				
نوع الإرتباط	إرتباط تابعي سالب (علاقة عكسية)	إرتباط سالب	إستقلال (عدم إرتباط)	إرتباط موجب	إرتباط تابعي موجب (علاقة طردية
الوضعية النسبية لمستقيمي الإنحدار	منطبقان		أحدهما عمودي و الأخر أفقي		منطبقان

تطبيق :

أحسب معامل الإرتباط ر في السلسلتين الإحصائيتين التاليتين:

x_i	12	5	9	7	3	6
y_i	1	3	5	0	1	2

الحل:

$$\lambda^2 = a_y(x) \times a_x(y)$$

$$a_y(x) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad , \quad a_x(y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$\lambda^2 = \frac{\left[\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right]^2}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\left[(y_i - \bar{y})^2\right]}$$

إذن:

$$\bar{x} = \frac{\sum_{i=1}^{6} x_i}{n} = \frac{42}{6} = 7 \quad , \quad \bar{y} = \frac{\sum_{i=1}^{6} y_i}{n} = \frac{12}{6} = 2$$

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
12	1	5	-1	-5	25	1
5	3	-2	1	-2	4	1
9	5	2	3	6	4	9
7	0	0	-2	0	0	4
3	1	-4	-1	4	16	1
6	2	-1	0	0	1	0
42	**12**	**0**	**0**	**3**	**50**	**16**

$$\lambda^2 = \frac{(3)^2}{(50)(16)} = \frac{9}{800} = 0{,}01125 \Rightarrow \boxed{\lambda = 0{,}106}$$