تمهيد:

من خلال ما تم تناوله سابقا، دائما ما كنا نتعرض لظواهر ذات المتغير الواحد وأدوات تحليلها، غير أنه كثيرا ما تصادفنا ظواهر ذات متغيرين أو أكثر تربطهما علاقة، قد تكون طردية؛ أي زيادة متغير أو نقصانه تؤدي إلى زيادة أو نقصان المتغير الآخر في نفس الاتجاه، كما قد تكون عكسية؛ أي زيادة متغير تؤدي إلى نقصان المتغير الآخر والعكس صحيح.

في هذه المحاضرة سوف يتم التطرق إلى:

- الارتباط؛
- الانحدار.

1. **الانحدار الخطي البسيط:** إن الغرض من استخدام تحليل الانحدار الخطي البسيط هو دراسة وتحليل أثر متغير كمي على متغير كمي آخر.

1.1 الانحدار الخطي البسيط التام: هنا تكون جميع نقاط شكل الانتشار على استقامة واحدة (استقامة تامة)، سواء كانت في الاتجاه الموجب أو في الاتجاه السالب.

تكون معادلة الإنحدار على الشكل التالي:

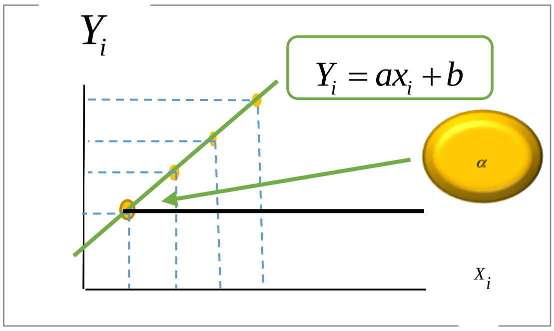
$$Y_i = ax_i + b$$

حيث a ميل دالة الانحدار، وb ثابت الدالة.

- ✓ A ميل الدالة؛ أي ظل الزاوية α حيث:

$$\alpha = \frac{\Delta Y}{\Delta X}$$

- ✓ b نقطة تقاطع خط الانحدار مع محور التراتيب.



ملاحظة هامة: لرسم شكل الانتشار لابد من تحديد المتغير التابع والذي نرمز له بالرمز (Y_i) ويمثل على محور التراتيب، طبعا المتغير التابع يتأثر بالمتغير المستقل والذي نرمز له (Xi) والذي يمثل على محور الفواصل.

2.1 الانحدار الخطي البسيط غير التام: في هذه الحالة نقاط شكل الانتشار لا تكون على استقامة تامة، ولكنها تأخذ إتجاها يمكن تقريبه من معادلة خط مستقيم، حيث أنه يستحيل إيجاد المعادلة الحقيقية والتي نفترضها كما يلي:

$y_i = ax_i + b + e_i$، حيث $ax_i + b$ هي معادلة خط مستقيم مضاف إليه المقدار e_i الذي يمثل قيمة البعد بين النقاط الحقيقية ومعادلة خط المستقيم. حيث أنه يستحيل إيجاد هذه المعادلة لذلك يتم تقريبها إلى المعادلة الخطية التالية:

$\hat{y}_i = \hat{a} x_i + \hat{b}$ حيث:

- $\hat{y}_i$ قيمة تقديرية؛
- $\hat{a}$ قيمة تقديرية لـ a؛

- $\hat{b}$ قيمة تقديرية لـ b.

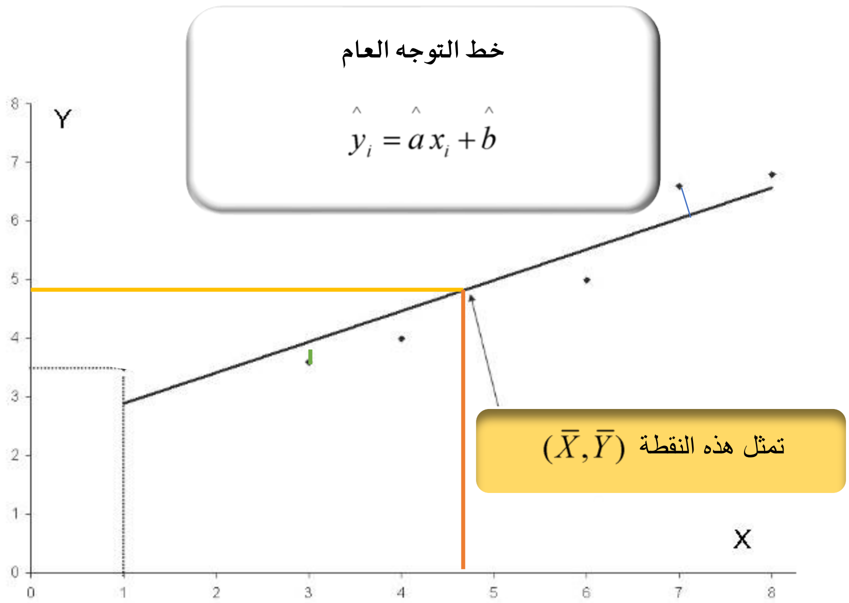
يمر الخط المستقيم من النقطة $(\bar{X},\bar{Y})$ ، حيث يمثل هذا المستقيم أقرب ما يمكن إلى جميع نقاط شكل الانتشار الحقيقية.

يتم إيجاد ثوابت الانحدار بما يسمى طريقة المربعات الصغرى والتي تعطي المعادلتين التقديريتين كما يلي:

$$\hat{a} = \frac{\sum_{i=1}^{n}(x_i-\bar{X})(y_i-\bar{Y})}{\sum_{i=1}^{n}(x_i-\bar{X})^2}$$

$$\hat{b} = \bar{Y} - \hat{a}\,\bar{X}$$

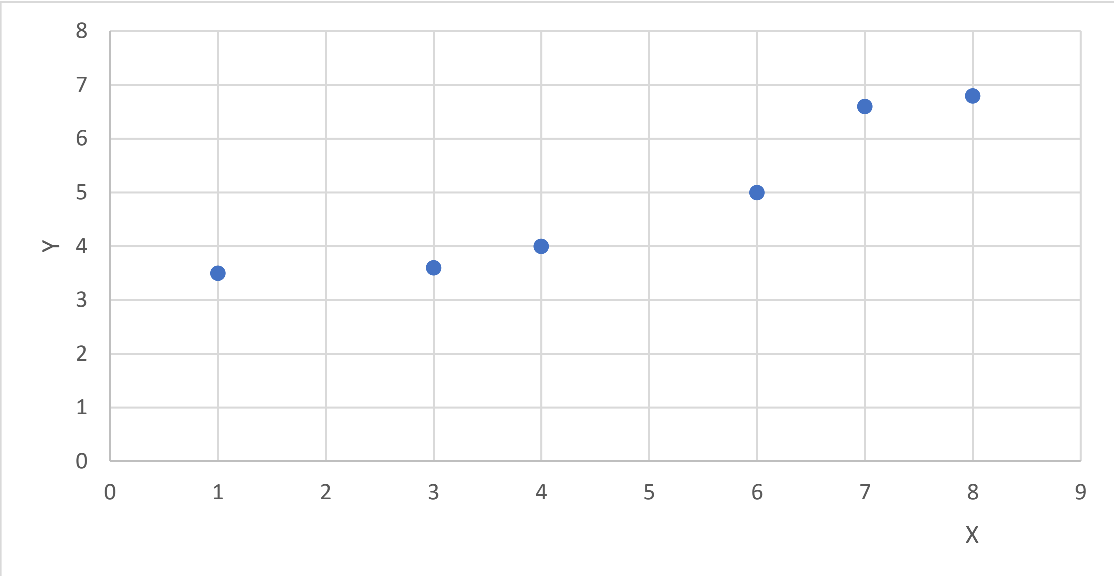
حيث أن: $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$ و $\bar{Y} = \frac{\sum_{i=1}^{n} y_i}{n}$



مثال 1: لتكن لدينا القيم التالية من الشكل (xi, yi)، حدد معادلة الانحدار :

$$S = \{(1,3.5)(3,3.6)(4,4)(6,5)(7,6.6)(8,6.8)\}$$

- شكل الانتشار:



- معادلة الانحدار من الشكل: $\hat{y}_i = \hat{a}\, x_i + \hat{b}$

	Xi	Yi	$x_i - \bar{X}$	$y_i - \bar{Y}$	$(x_i - \bar{X})(y_i - \bar{Y})$	$(x_i - \bar{X})^2$
	1	3.5	-3.83	-1.42	5.431	14.69
	3	3.6	-1.83	-1.32	2.414	3.36
	4	4	-0.83	-0.92	0.764	0.69
	6	5	1.17	0.08	0.097	1.36
	7	6.6	2.17	1.68	3.647	4.69
	8	6.8	3.17	1.88	5.964	10.03
$\sum$	**29**	**29.5**	**0**	**0.00**	**18.32**	**34.83**

$$\bar{X} = \frac{29}{6} = 4.83$$

$$\bar{Y} = \frac{29.5}{6} = 4.92$$

$$\hat{a} = \frac{18.32}{34.83} = 0.52$$

$$\hat{b} = 4.92 - 0.52(4.83) = 2.4$$

إذن معادلة الانحدار هي : $\hat{y}_i = 0.52x_i + 2.4$

2. معاملات الارتباط:

الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع وقوة العلاقة بين متغيرين.

نرمز له بالرمز r، حيث $-1 \le r \le +1$ فمن حيث نوع العلاقة نجد إما علاقة عكسية بين المتغيرين $(r \prec 0)$ أو علاقة طردية بينهما $(r \succ 0)$، في حين لما يكون $(r = 0)$ هنا لا توجد علاقة بين المتغيرين أصلا.

ارتباط طردي					ارتباط عكسي				
قوي جدا	قوي	متوسط	ضعيف	ضعيف جدا	ضعيف جدا	ضعيف	متوسط	قوي	قوي جدا
$[0.9,1[$	$[0.7,0.9[$	$[0.5,0.7[$	$[0.3,0.5[$	$[0,0.3[$	$[-0.3,0[$	$[-0.5,-0.3[$	$[-0.7,-0.5[$	$[-0.9,-0.7[$	$[-1,-0.9[$

r=0

ملاحظات هامة:

- لما $r = \pm 1$ نقول أنه يوجد ارتباط تام بين المتغيرين(كلي)؛
- لما $r = 0$ نقول لا يوجد ارتباط (انعدام العلاقة) بين المتغيرين؛
- $r \prec 0$ هناك علاقة عكسية بين المتغيرين؛
- $r \succ 0$ هناك علاقة طردية بين المتغيرين.

1.2. معامل الارتباط الخطي البسيط: يعرف بمعامل الارتباط لبيرسون وهو خاص بالمتغيرات الكمية ويحسب بالعلاقة التالية:

$r = \dfrac{\operatorname{cov}(x,y)}{\sigma_X . \sigma_y}$، حيث cov (x,y) هو التباين المشترك (التغاير) بين المتغيرين (x,y) حيث:

- $\operatorname{cov}(x,y) = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})$
- σ_X , σ_y الانحراف المعياري لكل من x وy.

يمكن اختصار علاقة حساب r كما يلي:

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\frac{\sum(x_i - \bar{X})^2}{n}} \times \sqrt{\frac{\sum(y_i - \bar{Y})^2}{n}}}$$

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum(x_i - \bar{X})^2} \times \sqrt{\sum(y_i - \bar{Y})^2}}$$

مثال 2: أحسب معامل الارتباط للمثال السابق (المثال 1).

	Xi	Yi	$x_i-\bar{X}$	$y_i-\bar{Y}$	$(x_i-\bar{X})(y_i-\bar{Y})$	$(x_i-\bar{X})^2$	$(y_i-\bar{Y})^2$
	1	3.5	-3.83	-1.42	5.431	14.69	2.01
	3	3.6	-1.83	-1.32	2.414	3.36	1.73
	4	4	-0.83	-0.92	0.764	0.69	0.84
	6	5	1.17	0.08	0.097	1.36	0.01
	7	6.6	2.17	1.68	3.647	4.69	2.83
	8	6.8	3.17	1.88	5.964	10.03	3.55
$\sum$	29	29.5	0	0.00	18.32	34.83	10.97

$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{X})(y_i-\bar{Y})}{\sqrt{\sum(x_i-\bar{X})^2}\times\sqrt{\sum(y_i-\bar{Y})^2}}=\frac{18.32}{\sqrt{34.83}\times\sqrt{10.97}}=\frac{18.32}{5.9\times3.31}\approx0.94$$

2.2. **معامل ارتباط الرتب**: إذا كانت متغيرات الدراسة متغيرات وصفية ترتيبية (كيفية)، فإنه يمكن حساب معامل ارتباط لسبيرمان يعتمد أساسا على رتب المتغيرين، يعبر عنه بالعلاقة التالية:

$r_{spearman}=1-\frac{6\sum_{i=1}^{n}d_i^2}{n(n^2-1)}$	• حيث d هي الفرق بين ترتيب x وy؛ • $d_i=R_X-R_Y$ • n عدد البيانات (x, y).

ملاحظة: يمكن كذلك استخدام صيغة معامل ارتباط الرتب في حساب **الارتباط بين متغيرين كميين**، حيث يتم استخدام رتب القيم التي يأخذها المتغير محل الدراسة.

مثال 3: الجدول التالي يوضح عدد سنوات التدخين X ونسبة تلف الرئة لعدد من المرضى في إحدى المستشفيات.

سنوات التدخين (X)	22	21	48	35	40	14
نسبة تلف الرئة (Y)	50	50	75	60	70	25

المطلوب:

- أحسب معامل سبيرمان للارتباط الرتب ثم فسره.

الحل:

❖ **معامل الارتباط سبيرمان:**

X (مرتب تصاعديا)	14	21	22	35	40	48
R (X)	1	2	3	4	5	6
Y (مرتب تصاعديا)	25	50	50	60	70	75
R(y)	1	2 2.5	3 2.5	4	5	6

نلاحظ أن قيمة Y للرتب 2 و 3 هي نفسها وتساوي 50، على هذا الأساس تأخذ نفس الرتبة؛ أي $\frac{2+3}{2}=2.5$

إذن: **R(y)= 1, 2.5, 2.5, 4, 5, 6**

Xi	Yi	R_X	R_Y	$d=R_X-R_Y$	d^2
22	50	3	2.5	0.5	0.25
21	50	2	2.5	-0.5	0.25
48	75	6	6	0	0
35	60	4	4	0	0
40	70	5	5	0	0
14	25	1	1	0	0
$\sum_{i=1}^{n=6}$					0.5

$$r_{spearman}=1-\frac{6\sum_{i=1}^{n}d_i^2}{n(n^2-1)}$$

$$r=1-\frac{6(0.5)}{6(36-1)}=1-\frac{3}{210}=0.98$$

إذن هناك ارتباط طردي قوي بين سنوات التدخين ونسبة التلف في الرئة.

ملاحظة: نستطيع حساب معامل الارتباط لبيرسون

مثال 4: تقوم إحدى المؤسسات بالتوظيف اعتمادا على امتحان كتابي ثم إجراء مقابلة شخصية. تدرس هذه المؤسسة إمكانية الاعتماد فقط على المقابلة الشخصية في التوظيف، من أجل ذلك أخذت عينة مكون من 8 أفراد ممن تقدموا للمقابلة سابقا، الجدول التالي يوضح لنا نتائج الامتحان الكتابي والمقابلة الشخصية.

علامة الامتحان الكتابي(X)	ممتاز	ضعيف	جيد	مقبول	جيد جدا	ممتاز	مقبول	جيد جدا
علامة المقابلة(Y)	جيد	جيد	جيد جدا	ضعيف	ضعيف	ممتاز	جيد	مقبول

السؤال: هل هناك علاقة بين علامة الامتحان الكتابي والمقابلة الشخصية؟، بما تنصح المؤسسة في عملية التوظيف.

X	ضعيف	مقبول	مقبول	جيد	جيد جدا	جيد جدا	ممتاز	ممتاز
R (X)	**1**	2 **2.5**	3 **2.5**	4	5 **5.5**	6 **5.5**	7 **7.5**	8 **7.5**
Y	ضعيف	ضعيف	مقبول	جيد	جيد	جيد	جيد جدا	ممتاز
R(Y)	1 **1.5**	2 **1.5**	3	4 **5**	5 **5**	6 **5**	7	8

Xi	Yi	R_X	R_Y	$d = R_X - R_Y$	d^2
ممتاز	جيد	7.5	5	2.5	6.25
ضعيف	جيد	1	5	-4	16
جيد	جيد جدا	4	7	-3	9
مقبول	ضعيف	2.5	1.5	1	1
جيد جدا	ضعيف	5.5	1.5	4	16
ممتاز	ممتاز	7.5	8	-0.5	0.25
مقبول	جيد	2.5	5	-2.5	6.25
جيد جدا	مقبول	5.5	3	2.5	6.25
$\sum_{i=1}^{n=8}$					**61**

$$r_{spearman} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

$$r = 1 - \frac{6(61)}{8(64-1)} = 1 - \frac{366}{504} = 0.27$$

نلاحظ وجود ارتباط ضعيف جدا بين علامة الامتحان الكتابي وعلامة المقابلة الشخصية، وعلى هذا الأساس ننصح المؤسسة أن لا تكتفي فقط على المقابلة الشخصية في عملية التوظيف.