Linear regression

٨ ـ ٢ : الانحدار الخطي

يعتبر العالم الانكليزي Francis Galton (1822-1911) أول من استخدم مفهوم الانحدار في التطبيقات البايولوجية بهدف اكتشاف بعض العلاقات بين بعض المتغيرات البايولوجية .

ويعرف الانحدار (أو تحليل الانحدار regression analysis) بشكل عام بانه مقياس رياضي لمتوسط العلاقة بين متغيرين أو أكثر بدلالة وحدات قياس المتغيرات المتمدة في العلاقة وغالباً ماتسمى العلاقات من هذا النوع بنماذج الانحدار . ان موضوع الانحدار ومن حيث التحليل يقسم الى قسمين رئيسين هما الانحدار الخطي والانحدار اللاخطي non-Linear regression .

ويختص الانحدار الخطي بتحليل العلاقة بين متغيرين أو أكثر يظهر كل منها بأس مساوٍ لواحد في العلاقة المعتمدة في التحليل في حين يختص الانحدار اللاخطي بتحليل العلاقة بين متغيرين أو أكثر تظهر فيها المتغيرات بأسس متباينة أو بهيئة صيغ لوغارتمية أو متعدد الحدود وغيرها من الاحوال المخالفة للانحدار الخطي . على سبيل المثال العلاقة :

$Y = a + bX$ علاقة خطية ظهر فيها كل من X ، Y بأس مساوٍ لواحد .

$Y = a + bX_1 + CX_2$ علاقة خطية ظهر فيها كل من Y , X_2 , X_1 بأس مساوٍ لواحد .

$Y = a + bX + CX^2$ علاقة لاخطية تعبر عن متعدد حدود من الدرجة الثانية

$Y = aX^b$ علاقة لاخطية تعبر عن دالة اسية مكافئة للعلاقة $LogY = Loga + bLogX$ التي يظهر فيها المتغيران بصيغة لوغارتمية .

$Y = aX_1^b . X_2^c$ علاقة لاخطية تعبر عن دالة اسية مكافئة للعلاقة $LogY = Loga + bLogX_1 + CLogX_2$ التي تظهر فيها المتغيرات بصيغة لوغارتمية .

$Y = a . b^x$ علاقة لاخطية تعبر عن دالة اسية مكافئة للعلاقة $LogY = Loga + XLogb$ التي يظهر فيها المتغير Y بصيغة لوغارتمية . وغيرها من العلاقات الاخرى .

وبهدف اعطاء صورة واضحة عن هذا الموضوع وعدم الخروج عن اختصاص هذا الكتاب فان اهتمامنا سوف يتركز في دراسة الانحدار الخطي لعلاقات تتضمن متغيرين أو ثلاثة متغيرات فقط .

٨ ـ ٢ ـ ١ : مفهوم الانحدار الخطي The concept of linear regression

ان موضوع تحليل الانحدار الخطي يهدف الى تقدير العلاقة الرياضية الخطية التي تربط متغيرين أو أكثر مع بعض . وفي تحليل الانحدار يوجد نوعين من المتغيرات هما المتغير التابع dependent variable والمتغير المستقل (أو مجموعة من المتغيرات المستقلة) independent variable . والمتغير التابع هو المتغير الذي يتأثر قيمه في حالة تغير قيم المتغير المستقل (المتغيرات المستقلة) ولا يتأثر (أو يؤثر) به . على سبيل المثال فان انفاق الاسرة الشهري يتأثر بدخلها الشهري في حين ان العكس غير منطقي . كذلك فان سرعة الانجاز لسباح تتأثر بمساحة كف يده وطولها في حين ان العكس غير منطقي . وغالباً مايسمى المتغير المستقل (المتغيرات المستقلة) المتغير المتنبأ به predictor أو المتغير (المتغيرات) التوضيحي explained variable ان موضوع الانحدار الخطي على صلة وثيقة بمفهوم الارتباط الخطي فالاول يهتم بتقدير العلاقة الخطية بين المتغيرات في حين ان الثاني يهتم بتقدير القيمة العددية للعلاقة نفسها .

ان العلاقات الخطية بين المتغيرات تنقسم بشكل عام الى قسمين رئيسين هما العلاقات المحددة deteministic relations والعلاقات العشوائية stochastic relations. والعلاقة المحددة تعني ان التغير الحاصل في متغير تابع هو نتيجة لتغير متغير مستقل (أو مجموعة متغيرات مستقلة) محدد (أو محددة) ولايوجد غيره (غيرها) . على سبيل المثال العلاقة بين المسافة والسرعة والزمن علاقة محددة بحيث يمكن ايجاد أي منها بدلالة الاخرين فالسرعة اللازمة لقطع مسافة محددة ماهي الا حاصل قسمة المسافة على الزمن المستغرق كذلك فان الزمن اللازم لقطع تلك المسافة ماهو الا حاصل قسمة المسافة على السرعة . وهذا يعني ان لكل قيمة من قيم المتغير المستقل يقابلها قيمة واحدة فقط للمتغير التابع . في حين ان العلاقة العشوائية تعني ان التغير الحاصل في متغير تابع ليس نتيجة لتغير متغير مستقل (أو مجموعة متغيرات مستقلة) فحسب بل الى جملة متغيرات اخرى مجهولة غير مسيطر عليها من قبل الباحث . على سبيل المثال الانفاق الشهري للاسرة لايتأثر فقط بدخلها الشهري وعدد افرادها فحسب بل هنالك جملة من المتغيرات الاخرى المؤثرة على الانفاق كالذوق ، العادات والتقاليد ، العلاقات الاجتماعية مع أسر اخرى ، وغيرها . وهذا يعني ان لكل قيمة من قيم المتغير المستقل (أو المتغيرات المستقلة) سوف لن يقابلها قيمة واحدة فقط للمتغير التابع وانما مجموعة من القيم التي كل منها يتأثر بواحد أو أكثر من المتغيرات المجهولة .

ان موضوع الانحدار يعالج بالاساس العلاقات العشوائية وليس العلاقات المحددة . وكما ذكرنا في اعلاه فان العلاقة الخطية بين المتغيرات تعني عملية ربط هذه المتغيرات بنموذج انحدار يتضح فيه المتغير التابع (المتأثر) والمتغير المستقل (المؤثر) أو مجموعة متغيرات مستقلة (المؤثرة) . وهذا يعني ان هنالك دالة رياضية تجمع هذه المتغيرات مع بعض . فاذا كان Y متغير تابع و X متغير مستقل فذلك يعني ان $Y = f(X)$. واذا كان هنالك متغيرين مستقلين فان $Y = f(X_1, X_2)$. ان الدالة الاولى يمكن صياغتها بشكل خطي أي $Y = a + bx$ ، حيث ان a,d ثوابت حقيقية (معالم الدالة) ، تعبر عن خط مستقيم في المستوى X.Y ، في حين ان الدالة الثانية يمكن صياغتها بشكل خطي ايضاً أي $Y = a + b_1X_1 + b_2X_2$ ، b_2, b_1, a ثوابت حقيقية (معالم الدالة) ، تعبر عن مستوى خطي في الفضاء الثلاثي X_2, X_1, Y . هذان الشكلان يعبران عن دالتين (علاقتين) محددتين ، وحتى نتمكن من صياغتها عشوائياً يجب ان يتم الاخذ بنظر الاعتبار جملة المتغيرات الاخرى المجهولة . فاذا رمزنا لهذه المجموعة بالرمز u عندئذٍ يمكن اعادة صياغة هاتين الدالتين بالشكل العشوائي التالي $Y = f(X, u) = a + bX + u$ ، $Y = f(X_1, X_2, u) = a + b_1X_1 + b_2X_2 + u$ أن ماسبق توضيحه في اعلاه ينطبق على العلاقات المتضمنة لاكثر من متغيرين مستقلين وسواء كانت علاقات خطية أم غير خطية .

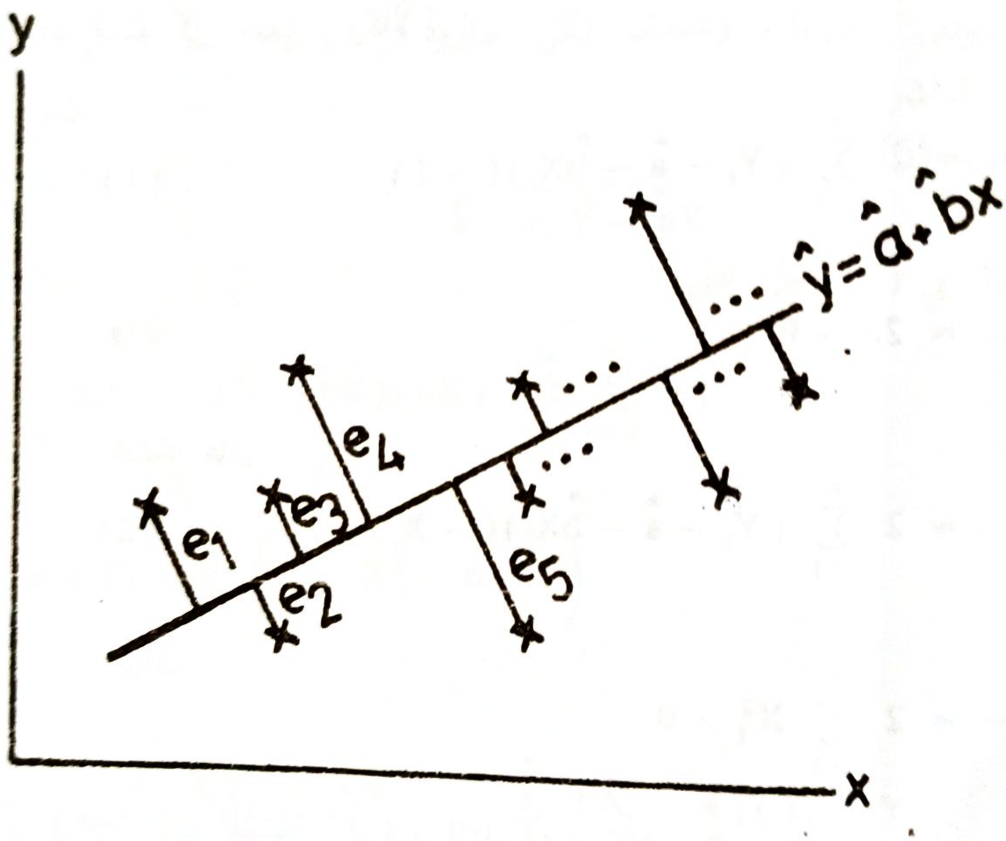
٨ ـ ٢ ـ ٢ : الانحدار الخطي البسيط Simple linear regression

يعرف الانحدار الخطي البسيط بأنه عملية تقدير العلاقة الخطية بين متغيرين احدهما مستقل والاخر تابع . ان مفهوم الانحدار الخطي البسيط يقترن بمفهوم الارتباط الخطي البسيط . ويهدف الانحدار الخطي البسيط الى تقدير قيم عددية لمعالم النموذج ، أي تقدير قيم عددية لكل من a ، b .

لاحظنا في الفقرة (٨ ـ ١ ـ ١) ان الشكل الانتشاري يبين بعض المعلومات المسبقة عما اذا كان المتغيرين مرتبطين أم لا وكذلك عما اذا كان هذا الارتباط موجبا ام سالباً . كذلك يمكن من خلال هذا الشكل معرفة طبيعة العلاقة بين المتغيرين فيما اذا كانت علاقة موجبة (زيادة في متغير تؤدي الى زيادة في الاخر والعكس صحيح) أو علاقة سالبة (زيادة في احدهما تؤدي الى انخفاض في الاخر والعكس صحيح) أو علاقة معدومة . وحيث اننا نتكلم عن علاقة عشوائية

تربط متغيرين فذلك يعني اننا بصدد ايجاد الخط المستقيم الذي يوضح هذه العلاقة أي ايجاد قيمة كل من a ، b التي تحدد شكل واتجاه هذا المستقيم . وهنالك طرق عديدة لتحديد ذلك اهمها هي طريقة المربعات الصغرى Least squares method التي يستند مبدئها الى ايجاد ذلك الخط المستقيم الذي يتخلل نقاط الشكل الانتشاري بالشكل الذي يجعل مجموع مربعات ابعاد النقاط عنه اقل مايمكن أي تحديد قيمة a ، b التي تجعل هذا المجموع أقل مايمكن . ان المستقيم الذي يتصف بهذه الميزة يسمى افضل مستقيم يعبر عن العلاقة بين المتغيرين .

افرض ان $Y = a + bX + u$ تعبر عن علاقة عشوائية (مجهولة) بين Y , X وافرض ان $\hat{Y} = \hat{a} + \hat{b}X$ يمثل الخط المستقيم الذي يعبر عن افضل علاقة تربط مابين Y , X كما في الشكل (٨ ــ ٤) .



شكل (٨ ــ ٤)

فاذا كانت العلاقة بين Y,X محددة فذلك يعني ان كافة النقاط سوف تقع على الخط المستقيم (أي ان أزواج النقاط (X_i , Y_i) تحقق المعادلة على سبيل المثال فان النقاط (0,2) ، (1,5) ، (2,8) تحقق المعادلة $Y = 2 + 3X$. لكن المشكلة هنا هي ان العلاقة عشوائية مما ادى ذلك الى تبعثر القيم (X_i , Y_i) في المستوى XY بشكل غير منتظم .

الان وعلى اساس عينة عشوائية من ازواج القيم (X_i, Y_i) قوامها n مفردة فان

$$e_i = Y_i - \hat{Y}_i \rightarrow e_i^2 = (Y_i - \hat{Y}_i)^2$$

$$\therefore \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{a} - \hat{b}X_i)^2 = f(\hat{a}, \hat{b})$$

والهدف الان هو ايجاد قيمة $\hat{b}, \hat{a}$ التي تجعل $\sum_{i=1}^{n} e_i^2$ اقل مايمكن . وهذا يعني اننا بصدد ايجاد النهاية الصغرى للدالة $f(\hat{a}, \hat{b})$. ان ذلك يتم من خلال ايجاد المشتقة الجزئية الاولى لهذه الدالة نسبة الى $\hat{a}$ والمشتقة الجزئية الاولى لهذه الدالة نسبة الى $\hat{b}$ ومساواة النواتج بالصفر والحل لغرض ايجاد قيمة كل منهما . وكالاتي :

ان

$$\frac{\partial f(\hat{a}, \hat{b})}{\partial \hat{a}} = 2 \sum_{i=1}^{n} (Y_i - \hat{a} - \hat{b}X_i)(-1) \qquad \ldots(1)$$

وان

$$\frac{\partial^2 f(\hat{a}, \hat{b})}{\partial \hat{a}^2} = 2n > 0.$$

كذلك فان

$$\frac{\partial f(\hat{a}, \hat{b})}{\partial \hat{b}} = 2 \sum_{i=1}^{n} (Y_i - \hat{a} - \hat{b}X_i)(-X_i) \qquad \ldots(2)$$

وان

$$\frac{\partial^2 f(\hat{a}, \hat{b})}{\partial \hat{b}^2} = 2 \sum_{i=1}^{n} X_i^2 > 0$$

وحيث ان المشتقة الجزئية الثانية في كلا حالتي الاشتقاق كانت موجبة فذلك مؤشر على ان الدالة $f(\hat{a}, \hat{b})$ تمتلك نهاية صغرى عند قيمتي $\hat{b}, \hat{a}$ الناتجتين من حل (1) ، (2) بعد مساواتها بالصفر .

عليه فإن

$$\sum_{i=1}^{n} (Y_i - \hat{a} - \hat{b}X_i) = 0$$

او ان

$$\sum_{i=1}^{n} Y_i = n\hat{a} + \hat{b} \sum_{i=1}^{n} X_i \qquad \therefore \bar{Y} = \hat{a} + \hat{b}\bar{X} \qquad \ldots(3)$$

كذلك فإن

فإذن

$$\sum_{i=1}^{n} (Y_i - \hat{a} - \hat{b}X_i) X_i = 0$$

$$\sum_{i=1}^{n} X_i Y_i = \hat{a} \sum_{i=1}^{n} X + \hat{b} \sum_{i=1}^{n} X_i^2$$

$$= \hat{a} n\bar{X} + \hat{b} \sum_{i=1}^{n} X_i^2 \quad \ldots (4)$$

ان المعادلتين (3) ، (4) سميان بالمعادلتين الطبيعيتين Normal equations كل منها بالمجهولين $\hat{b}, \hat{a}$. وعندئذ يمكن حلها انياً لايجاد $\hat{b}, \hat{a}$ أو عن طريق التعويض المباشر .

فمن المعادلة (3) لدنا

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

وبتعويضها في (4) نحصل على

$$\sum_{i=1}^{n} X_i Y_i = (\bar{Y} - \hat{b}\bar{X})(n\bar{X}) + \hat{b} \sum_{i=1}^{n} X_i^2$$

$$= n\bar{X}\bar{Y} + \hat{b} \left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \right)$$

وعندئذ :

$$\hat{b} = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} = \text{ص}$$

، يعني ان قيمة $\hat{b}$ ماهي الا حاصل قسمة التباين المشترك مابين X , Y ، تباين X ، وان $\hat{a} = \bar{Y} - \hat{b}\bar{X}$. هاتين القيمتين تحددان افضل خط مستقيم خلل الشكل الانتشاري معبراً عن العلاقة مابين Y , X . ويلاحظ ان كل من $\hat{b}, \hat{a}$ هي الان دالة بدلالة قياسات مفردات العينة (X_i, Y_i) .

ان $\hat{a}$ تمثل في الحقيقة مقطع خط الانحدار مع المحور الصادي اي مانعنيه قيمة Y التقديرية عندما $X = 0$. واذا كانت $\hat{a} = 0$ عندئذ يقال ان خط الانحدار يمر من نقطة الاصل (0,0) .
وان $\hat{b}$ تسمى معامل الانحدار regression coefficient أي مقياس يوضح مقدار تغير Y اذا ماتغيرت X بوحدة واحدة .

ويلاحظ أيضاً ان $\hat{b}$ قد تكون موجبة او سالبة تبعاً لقيمة S_{xy} ان كانت موجبة ام سالبة . فقيمة $\hat{b}$ الموجبة تعني ان العلاقة بين y,x موجبة (كما هي الحال بالنسبة لمعامل الارتباط البسيط) . والقيمة السالبة لها تعني ان العلاقة بين Y,X سالبة (عكسية) .
وغالباً مايقال لما سبق انه تم تقدير معادلة انحدار (Y/X) اي ايجاد معادلة انحدار فيها المتغير التابع هو (Y) والمتغير المستقل هو (X) .

مثال (١) : البيانات التالية تمثل الكمية المطلوبة (Y) من سلعة معينة وسعر الوحدة الواحدة منها (X) يطلب تقدير معادلة انحدار Y/X .

6	9	8	9	4	6	4	6	5	3	:Y
7	4	5	5	9	6	8	7	8	11	:X

الحل : نعمل الجدول التالي :

$X_i Y_i$	X_i^2	Y_i	X_i
33	121	3	11
40	64	5	8
42	49	6	7
32	64	4	8
36	36	6	6
36	81	4	9
45	25	9	5
40	25	8	5
36	16	9	4
42	49	6	7
382	530	60	70

واضح ان $\bar{Y} = 6 , \bar{X} = 7$

$$\therefore \hat{b} = \frac{382 - 10(7)(6)}{530 - 10(7^2)} = \frac{-38}{40} = -0.95$$

$$\therefore \hat{a} = 6 - (-0.95)(7) = 12.65$$

معادلة الانحدار المطلوبة هي

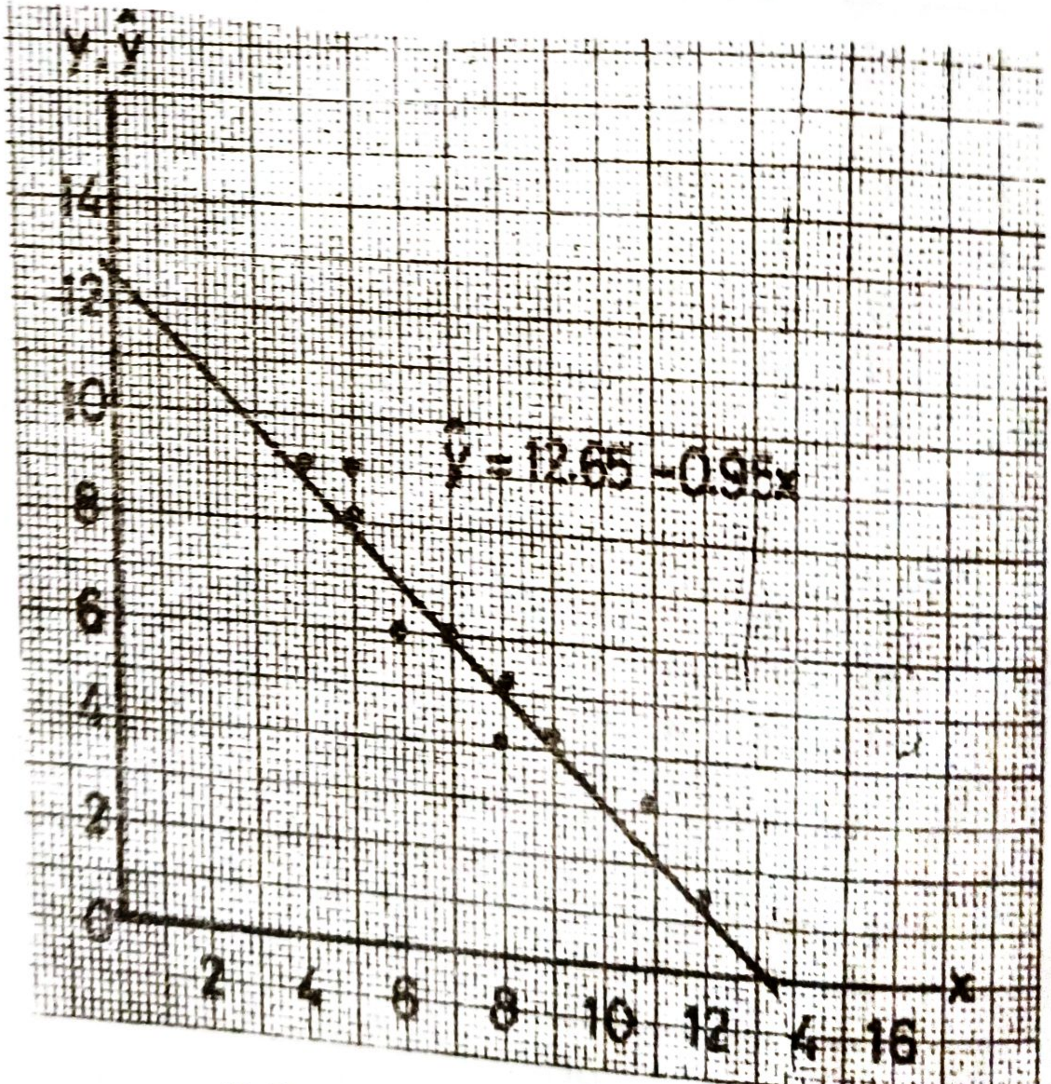
$$\hat{y} = 12.65 - 0.95X$$

لاحظ ان العلاقة بين Y, X سالبة (لماذا).

ولغرض رسم المعادلة اعلاه لختار ثلاثة قيم الى X (لرسم معادلة مستقيم نحتاج الى تعين ثلاثة نقاط في المستوي (XY) ونعوضها في المعادلة $\hat{Y}$ اي ان

X :	4 ,	8 ,	12
$\hat{Y}$:	8.85 ,	5.05 ,	1.25

وعليه فان رسم المعادلة والشكل الانتشاري معاً هو الموضح في الشكل (٨ ــ ٥).



شكل (٨ ــ ٥) : مخطط معادلة انحدار Y / X

ان احدى استخدامات معادلة الانحدار التقديرية هي لاغراض التنبؤ لقيمة Y اذا علمت قيمة الى X . فللمثال السابق اذا كان سعر الوحدة من هذه السلعة هو 13 فان الكمية التي يتوقع ان تطلب سوف تكون $\hat{Y} = 12.65 - 0.95 (13)$ أي 0.3 واذا كان سعر الوحدة من هذه السلعة هو 2 فان الكمية التي يتوقع ان تطلب سوف تكون $\hat{Y} = 12.65 - 0.95 (2) = 10.75$

مثال (٢) : البيانات التالية تمثل درجات 12 طالب في امتحان مادة الرياضيات (X) وامتحان مادة الاحصاء (Y) وان درجة الامتحان القصوى هي 10 درجات يطلب ايجاد معادلة انحدار X/Y, Y/X

1, 0, 6, 3, 6, 5, 10, 7, 8, 9, 3, 2, : X

0, 1, 4, 4, 6, 3, 9, 5, 7, 7, 2, 0, : Y

الحل : نعمل الجدول التالي :

$X_i Y_i$	Y_i^2	X_i^2	Y_i	X_i
0	0	4	0	2
6	4	9	2	3
63	49	81	7	9
56	49	64	7	8
35	25	49	5	7
90	81	100	9	10
15	9	25	3	5
36	36	36	6	6
12	16	9	4	3
24	16	36	4	6
0	1	0	1	0
0	0	1	0	1
337	286	414	48	60

ان $\bar{Y} = 4 , \bar{X} = 5$

١ ـ معادلة انحدار Y/X :

$$\hat{b} = \frac{337 - (12)(5)(4)}{414 - (12)(5^2)} = \frac{97}{114} = 0.85$$

$$\hat{a} = \bar{Y} - \hat{b}\bar{X} = 4 - (0.85)(5) = -0.25$$

$$\therefore \hat{Y} = -0.25 + 0.85X$$

لاحظ ان العلاقة موجبة (لماذا)

٢ ـ معادلة انحدار X/Y

$$\hat{b} = \frac{337 - (12)(4)(5)}{286 - (12)(4^2)} = \frac{97}{94} = 1.03$$

$$\therefore \hat{a} = \bar{X} - \hat{b}\bar{Y} = 5 - 1.03(4) = 0.88$$

$$\therefore \hat{X} = 0.88 + 1.03Y$$

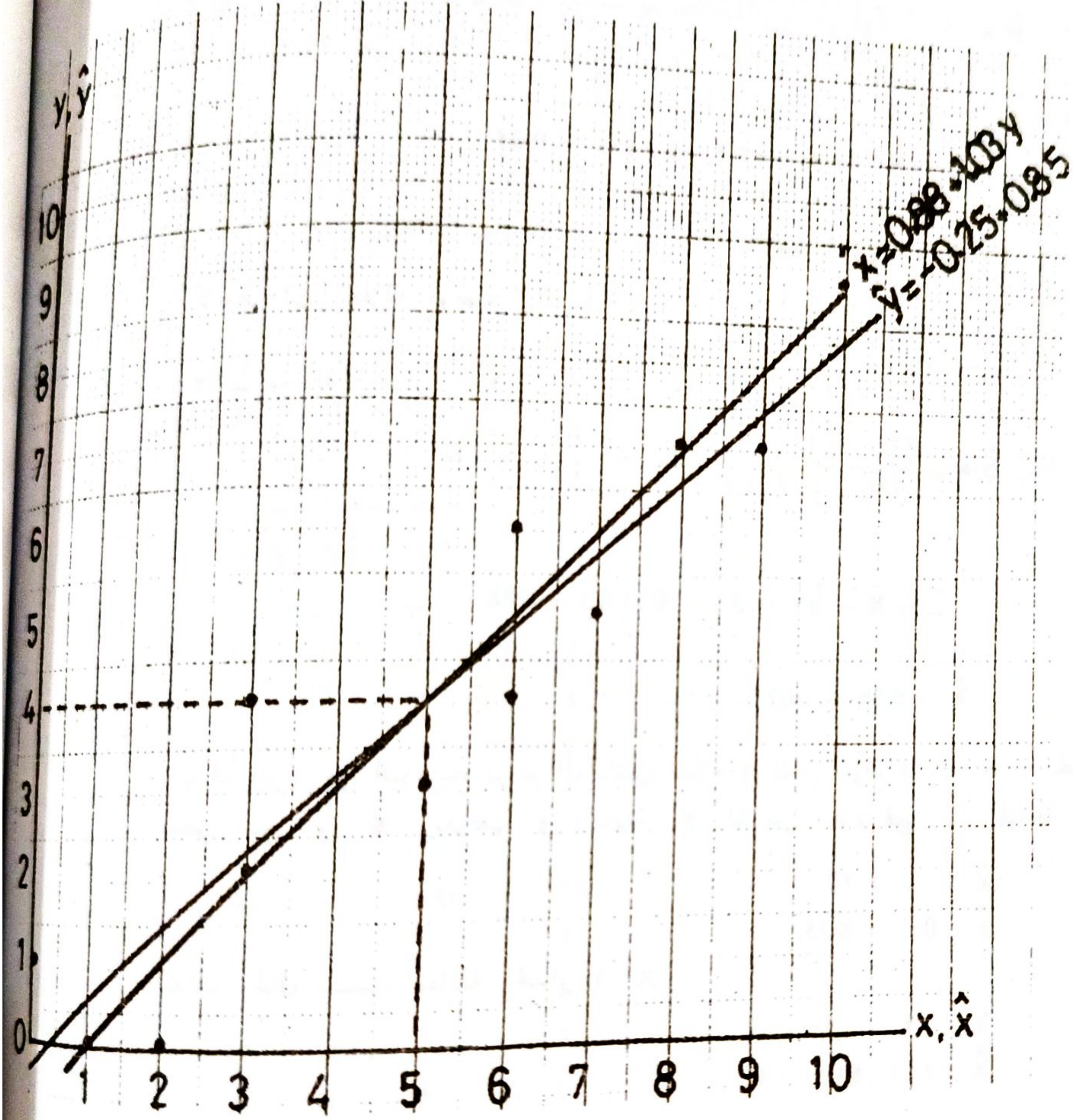
وبهدف المزيد من التوضيح نرسم المعادلتين اعلاه في نفس الموقع.لمعادلة Y/X نختار بعض القيم الى X ونعوضها في المعادلة $\hat{Y}$ لغرض ايجاد قيم $\hat{Y}$ القابلة .

X :	1	5	10
$\hat{Y}$:	0.6	4	8.25

كذلك الحال بالنسبة لمعادلة انحدار X Y

Y :	1	5	8
$\hat{X}$:	1.91	6.03	9.12

والآتي رسم المعادلتين معاً مع الشكل الانتشاري الموضحين في الشكل (٨ ـ ٦) :



شكل (٨ _ ٦) : مخطط انحدار Y / X ، X / Y

يلاحظ من الشكل اعلاه ان نقطه تقاطع انحدار Y/X مع انحدار X/Y هي النقطة ($\bar{X} = 5 , \bar{Y} = 4$).

كذلك يمكن استخدام المعادلتين المقدرتين للتنبؤ بقيمة احد المتغيرين اذا علمت قيمة الاخر . فمثلاً اذا كانت درجة طالب في الرياضيات هي 4 فمن المتوقع ان تكون درجته في الاحصاء هي $\hat{Y} = 3.15$ اي $\hat{Y} = -0.25 + (0.85)\ (4)$

كذلك اذا كانت درجة طالب في الاحصاء هي 8 فمن المتوقع ان تكون درجته في الرياضيات هي $9.12 = (8)\ (1.03) + 0.88 = \hat{X}$.

وفيا يلي بعض الملاحظات عن الانحدار الخطي البسيط :

١ ـ العلاقة بين معامل الانحدار ومعامل الارتباط البسيط .

يمكن حساب معامل الانحدار اذا علمت قيمة معامل الارتباط البسيط بين X , Y وفق الصيغة :

$$\hat{b} = \frac{S_y}{S_x} . r_{xy}$$

البرهان : لدينا

$$\hat{b} = \frac{S_{xy}}{S_x^2} , r_{xy} = \frac{S_{xy}}{S_x S_y}$$

وهذا يعني ان

$$S_x \hat{b} = \frac{S_{xy}}{S_x} = S_y r_{xy}$$

$$\therefore S_x \hat{b} = S_y r_{xy}$$

$$\therefore \hat{b} = \frac{S_y}{S_x} . r_{xy} , r_{xy} = \frac{S_x}{S_y} . \hat{b}$$

وحيث ان S_x/S_y ، S_y/S_x كميات موجبة دائماً لذا فان اشارة معامل الانحدار تتحدد من خلال اشارة معامل الارتباط والعكس صحيح دلالة على طبيعة العلاقة بين Y,X .

مثـــال (١) : اذا علمـــت ان $r_{xy} = 0.905$ وإن $S_x = 1.333$ ، $S_y = 2.211$.

جد معامل انحدار X/Y, Y/X .

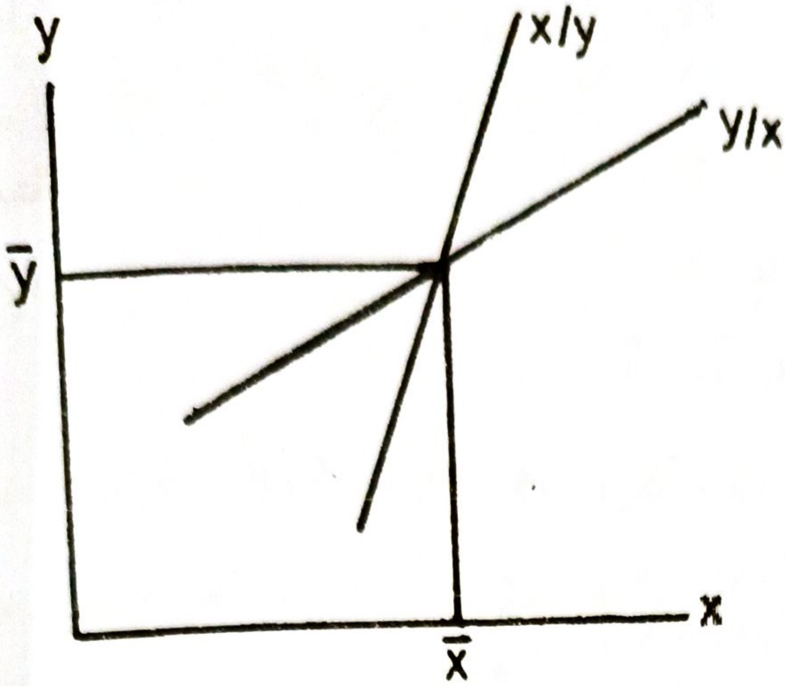
الحل

$$\text{coefficient of}(Y/X) = \frac{S_y}{S_x} r_{xy} = \frac{2.211}{1.333} \; 0.905$$

$$= 1.501$$

$$\text{coefficient of}(X/Y) = \frac{S_x}{S_y} r_{xy} = \frac{1.333}{2.211} . 0.905$$

$$= 0.546$$



شكل (٨ ـ ٨)

$$\sum_{i=1}^{n} Y_i = n\hat{a} + \hat{b} \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} (\hat{a} + \hat{b} X_i) = \sum_{i=1}^{n} \hat{Y}_i$$

كذلك فان

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (Y_i - \hat{Y}_i) = 0 \text{ وان } \bar{Y} = \bar{\hat{Y}}$$

٩ ـ ان مربع معامل الارتباط البسيط بين المتغيرين Y,X في انحدار Y/X يسمى معامل التحديد coefficient of determination ، وغالباً مايرمز له بالرمز R^2 ، وهو مقياس يوضح مقدار مايوضحه (يفسره) المتغير المستقل X من تذبذب (اضطراب) في المتغير التابع Y أي مانعنيه درجة مساهمة X في التغير الحاصل في Y .

وحيث ان $-1 \le r_{xy} \le 1$ فان $0 \le R^2 = r_{xy}^2 \le 1$ وهذا يعني ان قيم هذا المقياس تتراوح بين الصفر والواحد .

وكما ذكرنا سابقاً فان الانحدار يعالج بالأساس العلاقات العشوائية وان سبب تغير Y ليس بسبب تغير X فقط وانما بجملة عوامل اخرى (متغيرات اخرى) غير مسيطر عليها. وعليه فاذا كان R^2 مقياس يوضح مقدار مايفسره X من تذبذب في Y فان $(1 - R^2)$ سوف يمثل مقياس يوضح مقدار ماتفسره جملة المتغيرات الاخرى، التي رمزنا لها بـ u ، من التذبذب الحاصل في Y الذي لايعزى الى X .

وكلما ارتفعت قيمة R^2 فذلك مؤشر على ان X ذا تأثير كبير على Y والعكس صحيح أيضاً . كذلك اذا كانت قيمة R^2 مساوية للصفر فذلك يعني ان X لايمتلك اي تأثير على Y وانما هنالك متغيرات اخرى هي التي تفسر الاضطراب في Y . في حين اذا كانت قيمة R^2 مساوية لواحد فذلك يعني ان X هو المتغير الوحيد الذي يفسر سبب الاضطراب في Y .

ويمكن حساب قيمة معامل التحديد وفق احدى الصيغ التالية التي هي في الحقيقية مشتقة الواحدة من الاخرى :

$$^2 = r_{xy}^2 \qquad \ldots (1)$$

$$1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} y_i^2}, \; y_i = Y_i - \bar{Y} \qquad \ldots (2)$$

$$= \; ^2{}_{/x} \cdot \frac{S_x^2}{S_y^2} \qquad \ldots (3)$$

$$= \frac{\sum_{i=1} \hat{y}_i^2}{\sum_{i=1} y_i^2}, \quad \hat{y}_i = (\hat{Y}_i - \bar{Y}) \qquad \ldots (4)$$

البرهان :

لدينا

$$e_i = Y_i - \hat{Y}_i \;\rightarrow\; Y_i = \hat{Y}_i + e_i \qquad \therefore\; Y_i - \bar{Y} - \hat{Y}_i - \bar{Y} + e$$

علماً ان $\bar{Y} = \bar{\hat{Y}}$

افرض ان : $y_i = Y_i - \bar{Y}, \hat{y}_i = \hat{Y}_i - \bar{Y}$

فاذن

$$y_i = \hat{y}_i + e_i \;\rightarrow\; \sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} (\hat{y}_i + e_i)^2$$

$$\therefore \sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} \hat{y}_i^2 + \sum_{i=1}^{n} e_i^2 + 2\sum_{i=1}^{n} \hat{y}_i e_i$$

$$= 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} y_i^2}$$

مثال (٦) : لبيانات المثال (١) فان

$$R^2 = r_{xy}^2 = (0.905)^2 = 0.82$$

$$= \hat{b}_{y/x}^2 \cdot \frac{S_x^2}{S_y^2} = (1.501)^2 \cdot \frac{(1.333)^2}{(2.211)^2} = 0.82$$

$$= \hat{b}_{x/y}^2 \cdot \frac{S_y^2}{S_x^2} = (0.546)^2 \cdot \frac{(2.211)^2}{(1.333)^2} = 0.82$$

ففي انحدار Y/X فان $R^2 = 82\%$ تعني ان المتغير X يوضح 82% من التذبذب الحاصل في Y والمتبقي 18% يمثل التذبذب الموضح من قبل المتغيرات الاخرى غير المسيطر عليها .

١٠ - الخطأ المعياري لتقدير معادلة انحدار Y/X (الانحراف المعياري للخطأ) :

يعتبر الخطأ المعياري (الانحراف المعياري) لتقدير معادلة انحدار Y/X احد المؤشرات المهمة التي تبين جودة توفيق معادلة انحدار Y/X، اي مانعنيه مقدار دقة تمثيل معادلة الانحدار للعلاقة بين المتغيرين X، Y على ضوء البيانات المتاحة عن كل منهما . وتبرز اهمية هذا المؤشر عند حسابنا لفترة ثقة معالم الانحدار وكذلك الاختبارات الاحصائية المتعلقة بها التي سنأتي الى ذكرها في الفصول القادمة . وغالباً مايرمز لهذا المؤشر بالرمز $S_{y/x}$. كذلك يستفاد من هذا المؤشر لدى اجراء المقارنة بين معادلتي انحدار او اكثر حول نفس الظاهرتين في دراستين مستقلتين او اكثر .

وكلما انخفضت قيمة $S_{y/x}$ فذلك مؤشر على دقة التوفيق .

ويمكن حساب قيمة $S_{y/x}$ وفق الصيغة التالية :

$$S_{y/x} = \sqrt{\sum_{i=1}^{n} e_i^2 / n-2} \quad , e_i = Y_i - \hat{Y}_i$$

وهذا يعني ان الخطأ المعياري ماهو الا الجذر التربيعي الموجب لتباين الخطأ .

كذلك فان $S_{y/x} > 0$.

مثال (٧) : اذا علمت ان R^2 في انحدار Y / X على اساس عينة قوامها 10 مفردات كانت مساوية الى 0.82 وأن $S_y^2 = 4.89$. جد $S_{y/x}$

الحل :

$$R^2 = 0.82 \rightarrow 1 - R^2 = 0.18$$

حيث ان

$$1 - R^2 = \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} y_i^2} = 0.18$$

لكن

وان

$$\sum_{i=1}^{n} y_i^2 = nS_y^2 = 10(4.89) = 48.9$$

$$\therefore \sum_{i=1}^{n} e_i^2 = (48.9)(0.18) = 8.802$$

$$\therefore S_{y/x} = \sqrt{\frac{8.802}{8}} = 1.049$$

مثال (٨) : البيانات التالية تمثل الكمية المعروضة (Y) من سلعة معينة وسعر الوحدة الواحدة منها (X) .

X : 2 , 4 , 3 , 5 , 3 , 7 , 4

Y : 4 , 6 , 5 , 8 , 4 , 10 , 5

يطلب اجراء مايلي :

أ ـ تقدير دالة العرض لهذه السلعة .

ب ـ تقدير دالة السعر لهذه السلعة .

جـ ـ رسم دالة العرض ودالة السعر المقدرتين ثم تحقق من الرسم ان نقطة تقاطع الدالتين هي $(\bar{X}, \bar{Y})$.

د ـ جد معامل الارتباط البسيط بين Y , X .

هـ ـ بين ان $\hat{X} = \bar{X}, \hat{Y} = \bar{Y}$

و ـ بين ان $\Sigma e_i = 0$ في انحدار Y / X

ز ـ جد معامل التحديد في انحدار Y / X ثم فسر معناه .

ح ـ جد الخطأ المعياري لتقدير انحدار Y / X.

الحل : نعمل الجدول التالي : الذي يبين الحسابات للفقرات المطلوب اجراؤها

e_i^2	e_i	$\hat{X}_i$	$\hat{Y}_i$	XY	Y^2	X^2	Y	X
4	3	2	1					
0·3906	0·6250	2·6	3·3750	8	16	4	4	2
0·0000	0·0000	4·0	6·0000	24	36	16	6	4
0·0977	0·3125	3·3	4·6875	15	25	9	5	3
0·4727	0·6875	5·4	7·3125	40	64	25	8	5
0·4727	− 0·6875	2·6	4·6875	12	16	9	4	3
0·0039	0·0625	6·8	9·9375	70	100	49	10	7
1·0000	− 1·0000	3·3	6·0000	20	25	16	5	4
2·4376	0·0000	28	42	189	282	128	42	28

ان $\bar{Y} = 6, \bar{X} = 4$

أ ـ

$$\hat{b}_{y/x} = \frac{189 - 7(4)(6)}{128 - 7(4^2)} = 1{\cdot}3125$$

$$\therefore \hat{a}_{y/x} = 6 - (1{\cdot}3125)(4) = 0{\cdot}75$$

∴ دالة العرض التقديرية هي

$$\hat{Y} = 0{\cdot}75 + 1{\cdot}3125\, X$$

ب ـ

$$\hat{b}_{x/y} = \frac{189 - 7(4)(6)}{282 - 7(6^2)} = 0{\cdot}7$$

$$\therefore \hat{a}_{x/y} = 4 - (0{\cdot}7)(6) = -\,0{\cdot}2$$

∴ دالة السعر التقديرية هي

$$\hat{X} = -\,0{\cdot}2 + 0{\cdot}7\, Y$$

جـ ـ لغرض رسم المعادلتين نختار قيم للمتغير المستقل في كل حالة ونعوضها في المعادلة التقديرية لغرض الحصول على قيم تقديرية للمتغير التابع وكالاتي:

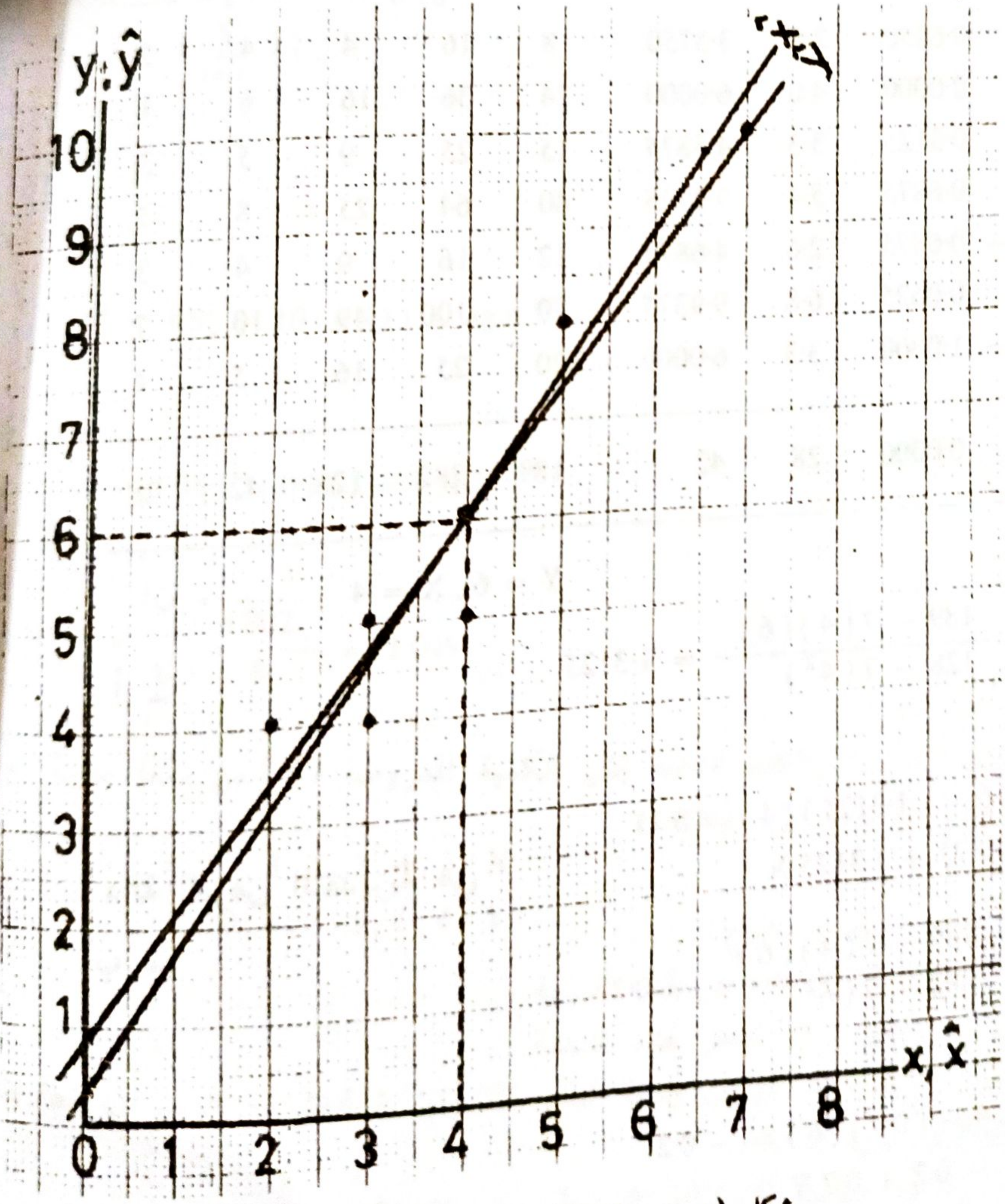
لمعادلة انحدار Y / X

X :	0	4	8
$\hat{Y}$:	0.75	6	11.25

لمعادلة انحدار X / Y

Y :	1	6	8
$\hat{X}$ =	0.5	4	5.4

ويتم تعين هذه النقاط ورسم المعادلتين المطلوبتين وكما هو موضح في الشكل (٨-٩):



شكل (٨ ـ ٩): مخطط انحدار Y/X وانحدار X/Y

لاحظ من الرسم ان نقطة تقاطع المستقيمين هي النقطة التي احداثيها السيني هو $\bar{X} = 4$ والصادي هو $\bar{Y} = 6$.

د ـ لغرض ايجاد معامل الارتباط فان ذلك يتم بسهولة من خلال ايجاد الوسط الهندسي مابين معامل انحدار Y/X ومعامل انحدار X/Y .

$$\therefore r_{xy} = \sqrt{\hat{b}_{y/x} \cdot \hat{b}_{x/y}} = \sqrt{(1\cdot3125)(0\cdot7)} = 0\cdot957$$

هـ ـ ان المطلوب واضح من الجدول السابق حيث ان $\Sigma \hat{X} = \Sigma X$ ، $\Sigma \hat{Y} = \Sigma Y$ لاحظ الحقلين 1 ، 2 .

و ـ لاحظ مجموع الحقل (3) في الجدول السابق .

ز ـ ان معامل التحديد في Y/X ماهو الا مربع الارتباط البسيط بين Y, X أي أن $R^2 = (0\cdot957)^2 = 0\cdot916$ وهذا يعني ان التغير في السعر يفسر حوالي 91.6% من التذبذب الحاصل في الكمية المعروضة من هذه السلعة .

ح ـ ان الخطأ المعياري لتقدير انحدار Y/X هو

$$S_{y/x} = \sqrt{\frac{\Sigma e_i^2}{n-2}} = \sqrt{\frac{2\cdot4376}{5}} = 0\cdot6982$$

حيث ان Σe_i^2 يمثل مجموع الحقل (4) في الجدول اعلاه .

٨ ـ ٢ ـ ٣ : الانحدار الخطي المتعدد (حالة متغيرين مستقلين) Multiple linear regression (two independent variables)

يعرف الانحدار الخطي المتعدد بانه عملية تقدير العلاقة الخطية بين عدة متغيرات احدها متغير تابع والباقي متغيرات مستقلة يعتقد انها تؤثر في المتغير التابع . وسوف يتركز اهتمامنا هنا بدراسة حالة متغيرين مستقلين فقط . ان مفهوم الانحدار الخطي المتعدد يقترن بمفهوم الارتباط الجزئي والارتباط المتعدد الذين سبق دراستها في (٨ـ١ـ٢) ، (٨ـ١ـ٣) من هذا الفصل .

افرض ان $Y = a + b_1X_1 + b_2X_2 + u$ دالة تعبر عن علاقة عشوائية بين المتغيرين المستقلين X_2, X_1 والمتغير التابع Y ليكن $\hat{Y} = \hat{a} + \hat{b}_1X_1 + \hat{b}_2X_2$ يمثل افضل مستوى انحدار في الفضاء الثلاثي الابعاد (Y, X_1, X_2) الذي يعبر عن العلاقة مابين هذه المتغيرات الثلاث .

الان وعلى اساس عينة عشوائية من المفردات قوامها n مفردة تم الحصول على بيانات كل متغير من هذه المتغيرات من كل مفردة عندئذ فان الفرق مابين قيم Y_i الحقيقية وقيم هذا المتغير التقديرية $\hat{Y}_i$ هو

$$e_i = Y_i - \hat{Y}_i , i = 1 , 2 , \ldots , n$$

وعندئذ فان :

$$f = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

ان هدفنا الان الحصول على قيم عددية الى كل من $\hat{b}_2 , \hat{b}_1 , \hat{a}$ التي تجعل f أقل مايمكن (أي استخدام مبدأ المربعات الصغرى). ومن خلال التعويض عن $\hat{Y}_i$ بما يساويها في f واجراء التفاضل الجزئي للدالة f نسبة لكل من $\hat{b}_2 , \hat{b}_1 , \hat{a}$ وجعل المشتقات الجزئية الثلاث مساوية للصفر نحصل على المعادلات الثلاث التالية التي تسمى بالمعادلات الطبيعية :

$$\sum_{i=1}^{n} Y_i = n\hat{a} + \hat{b}_1 \sum_{i=1}^{n} X_{1i} + \hat{b}_2 \sum_{i=1}^{n} X_{2i} \quad \ldots (1)$$

$$\sum_{i=1}^{n} X_{1i} Y_i = \hat{a} \sum_{i=1}^{n} X_{1i} + \hat{b}_1 \sum_{i=1}^{n} X_{1i}^2 + \hat{b}_2 \sum_{i=1}^{n} X_{1i} X_{2i} \quad \ldots (2)$$

$$\sum_{i=1}^{n} X_{2i} Y_i = \hat{a} \sum_{i=1}^{n} X_{2i} + \hat{b}_1 \sum_{i=1}^{n} X_{1i} X_{2i} + \hat{b}_2 \sum_{i=1}^{n} X_{2i}^2 \quad \ldots (3)$$

واضح ان هنالك ثلاث معادلات بثلاث مجاهيل هي $\hat{b}_2 , \hat{b}_1 , \hat{a}$ ؛ بحل هذه المعادلات انياً أو بالتعويض أو أي طريقة اخرى ممكنة نحصل على مايلي :

$$\hat{a} = \bar{Y} - \hat{b}_1 \bar{X}_1 - \hat{b}_2 \bar{X}_2$$

$$\hat{b}_1 = \frac{\left(\sum_{i=1}^{n} x_{1i} y_i \right) \left(\sum_{i=1}^{n} x_{2i}^2 \right) - \left(\sum_{i=1}^{n} x_{2i} y_i \right) \left(\sum_{i=1}^{n} x_{1i} x_{2i} \right)}{\left(\sum_{i=1}^{n} x_{1i}^2 \right) \left(\sum_{i=1}^{n} x_{2i}^2 \right) - \left(\sum_{i=1}^{n} x_{1i} x_{2i} \right)^2}$$

$$\hat{b}_2 = \frac{\left(\sum_{i=1}^{n} x_{2i} y_i\right)\left(\sum_{i=1}^{n} x_{1i}^2\right) - \left(\sum_{i=1}^{n} x_{1i} y_i\right)\left(\sum_{i=1}^{n} x_{1i} x_{2i}\right)}{\left(\sum_{i=1}^{n} x_{1i}^2\right)\left(\sum_{i=1}^{n} x_{2i}^2\right) - \left(\sum_{i=1}^{n} x_{1i} x_{2i}\right)^2}$$

حيث ان :

$$y_i = Y_i - \bar{Y} \,,\, x_{1i} = X_{1i} - \bar{X}_1 \,,\, x_{2i} = X_{2i} - \bar{X}_2$$

واضح مما تقدم ان كل من $\hat{b}_2 , \hat{b}_1 , \hat{a}$ ، دوال بدلالة قياسات العينة Y_i , X_{1i} , X_{2i} .

ان $\hat{a}$ في الحقيقة تمثل المقطع الصادي أي مقطع مستوى الانحدار مع المحور Y عندما $X_1 = X_2 = 0$. وان $\hat{b}_1$ تعني معامل انحدار Y على X_1 بثبوت اثر X_2 أي $\hat{b}_{Y/X_1 \cdot X_2}$ وان $\hat{b}_2$ تعني معامل انحدار Y على X_2 بثبوت اثر X_1 أي $\hat{b}_{Y/X_2 \cdot X_1}$. وهذا يعني ان $\hat{b}_1$ تمثل مقدار تغير Y عند تغير X_1 بوحدة واحدة عند ثبات X_2 وان $\hat{b}_2$ تمثل مقدار تغير Y عند تغير X_2 بوحدة واحدة عند ثبات X_1 . إن القيمة السالبة لاي من هذين المعاملين تعني وجود علاقة سالبة (عكسية) بين ذلك المتغير مع Y في حين ان القيمة الموجبة لاحدهما تعني وجود علاقة موجبة (طردية) بين ذلك المتغير مع Y .

مثال (١) : البيانات التالية تمثل الكمية المطلوبة من سلعة معينة (Y) ، سعر الوحدة من هذه السلعة (X_1) ، الدخل الشهري (X_2) لعينة قوامها 10 . يطلب ايجاد معادلة الانحدار التي تعبر عن العلاقة مابين هذه المتغيرات .

Y : 3 , 5 , 6 , 4 , 6 , 4 , 9 , 8 , 9 , 6

X_1 : 11 , 8 , 7 , 8 , 6 , 9 , 5 , 5 , 4 , 7

X_2 : 100 , 120 , 130 , 100 , 120 , 110 , 140 , 130 , 140 , 110

الحل : نعمل الجدول التالي الذي يوضح الحسابات المطلوبه :

x_1x_2	yx_2	yx_1	x_2^2	x_1^2	y^2	x_2	x_1	y	X_2	X_1	Y
− 80	60	− 12	400	16	9	− 20	4	− 3	100	11	3
0	0	− 1	0	1	1	0	1	− 1	120	8	5
0	0	0	100	0	0	10	0	0	130	7	6
− 20	40	− 2	400	1	4	− 20	1	− 2	100	8	4
0	0	0	0	1	0	0	− 1	0	120	6	6
− 20	20	− 4	100	4	4	− 10	2	− 2	110	9	4
− 40	60	− 6	400	4	9	20	− 2	3	140	5	9
− 20	20	− 4	100	4	4	10	− 2	2	130	5	8
− 60	60	− 9	400	9	9	20	− 3	3	140	4	9
0	0	0	100	0	0	− 10	0	0	110	7	6
− 240	260	− 38	2000	40	40	0	0	0	1200	70	60

$$\bar{X}_2 = 120\,,\quad \bar{X}_1 = 7\quad ,\bar{Y} = 6$$

$$\therefore \hat{b}_1 = \frac{(-38)(2000) - (260)(-240)}{(40)(2000) - (-240)^2} = \frac{-13600}{22400}$$

$$= -0.607$$

$$\hat{b}_2 = \frac{(260)(40) - (-38)(-240)}{(40)(2000) - (-240)^2} = \frac{1280}{22400}$$

$$= 0.057$$

$$\therefore \hat{a} = 6 - (-0.607)(7) - (0.057)(120) = 3.409$$

فاذن معادلة الانحدار المطلوبة هي

$$\hat{Y}_i = 3.409 - 0.607X_{1i} + 0.057X_{2i}$$

واضح من هذه المعادلة انه اذا كان $X_1 = 10$, $X_2 = 105$ فان $Y = 3.324$

واذا كان $X_1 = 3$, $X_2 = 150$ فان $Y = 10.138$

وفيا يلي بعض الملاحظات عن الانحدار الخطي المتعدد .

١ ـ ان مجموع قيم المتغير التابع Y الحقيقة مساوٍ لمجموع قيم هذا المتغير التقديرية اي ان

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \hat{Y}_i$$

وهذا يعني ان $\sum_{i=1}^{n} e_i = 0$. ونترك برهنة ذلك للقارئ .

٢ ـ ان حاصل ضرب معامل انحدار $Y / X_1 . X_2$ في معامل انحدار $X_1 / Y . X_2$ ماهو الا مربع معامل الارتباط الجزئي بين Y , X_1 عند استبعاد اثر X_2 . اي ان

$$r^2_{yx_1 \cdot x_2} = \hat{b}_{y/x_1 \cdot x_2} \cdot \hat{b}_{x_1/y \cdot x_2}$$

٣ ـ ان مربع معامل الارتباط المتعدد بين Y مع X_2 , X_1 يسمى بمعامل التحديد Coefficient of determination ويرمز له بالرمز $R^2_{y \cdot x_1 x_2}$ او للسهولة R^2 .. وهو مقياس يوضح مقدار مايفسره المتغيرين X_2 , X_1 المستقلين من تذبذب (اضطراب) في المتغير التابع Y اي مانعنيه درجة مساهمة X_2 , X_1 في التغير الحاصل في Y . ان مفهوم هذا المعامل هنا لايختلف بشيء عن مفهومه الذي سبق توضيحه في الملاحظة (٩) من الفقرة السابقة سوى انه هنا يتم التعامل مع متغيرين مستقلين وليس متغير واحد .

ويمكن حساب قيمة هذا المعامل بأحدى الطرق التالية التي كل منها تلائم المعلومات المتوفرة للحالة قيد الدراسة :

$$R^2_{y \cdot x_1 x_2} = \frac{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y}_i)^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2} \quad \ldots (1)$$

$$= 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} y_i^2} \quad \ldots (2)$$

$$= \frac{r^2_{yx_1} + r^2_{yx_2} - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r^2_{x_1x_2}} \qquad \ldots (3)$$

$$= \frac{\hat{b}_1 \sum_{i=1}^{n} x_{1i}y_i + \hat{b}_2 \sum_{i=1}^{n} x_{2i}y_i}{\sum_{i=1}^{n} y_i^2} \qquad \ldots (4)$$

مثال (٢) : بتطبيق الصيغة (4) على المعلومات المتوفرة في المثال (١) نلاحظ ان

$$R^2 = \frac{(-0.607)(-38) + (0.057)(260)}{40} = \frac{37.886}{40}$$

$$= 0.947$$

وهذا يعني ان المتغيرين X_2 , X_1 يفسران حوالي 94.7% من التذبذب الحاصل في المتغير Y والمتبقي البالغ 5.3% يعود لجملة المتغيرات الاخرى المؤثرة على Y الغير مسيطر عليها .

٤ ـ الخطأ المعياري لتقدير معادلة انحدار $Y / X_1 , X_2$

ان مفهوم الخطأ المعياري هنا لايختلف اطلاقاً عن مفهومه الذي سبق توضيحه في الملاحظة (١٠) من الفقرة السابقة . ويتم حساب الخطأ المعياري في حالة وجود متغيرين مستقلين وفق الصيغة التالية :

$$S_{y/x_1x_2} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-3}}$$

مثال (٣) : من المعلومات المتوفرة في المثال الثاني هي :

$$R^2 = 0.947 \quad , \sum_{i=1}^{n} y_i^2 = 40 \, , n = 10$$

وان

$$R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} y_i^2} \rightarrow 0.053 = \frac{\Sigma e_i^2}{40}$$

$$\therefore \Sigma e_i^2 = 2{\cdot}12$$

$$\therefore S_{y/x_1x_2} = \sqrt{\frac{2{\cdot}12}{7}} = 0{\cdot}55$$

٨ ـ ٢ ـ ٤ : مقارنة تطبيقية بين الانحدار الخطي البسيط والمتعدد .

ان دراسة العلاقة بين متغيرين احدهما تابع والاخر مستقل مرتبطين بعلاقة عشوائية تبدو من الناحية التطبيقية مسألة غير واقعية حيث انه لايوجد متغير يتأثر بمتغير واحد فقط وانما بجملة متغيرات اخرى بعضاً منها ممكن التحديد والقياس وان دراستنا للانحدار الخطي البسيط لاتعدو سوى عملية توضيح موضوع الانحدار من الناحية النظرية . كذلك فانه كلما ازداد عدد المتغيرات المستقلة فذلك يعني اضعاف في اهمية المتغير العشوائي u بسبب سحب عدد من مكوناته وتشخيصها في نموذج الانحدار وهذه العملية تؤثر في تقديرات معالم النموذج المقدر وفي المؤشرات المحتسبة له . وفيما يلي مقارنه تطبيقية للنتائج التي تم الحصول عليها من المثال (١) الوارد في الفقرة (٨ ـ ٢ ـ ٣) بعد تضمين النموذج المتغير X_2 (الدخل الشهري) . وان اسلوب المقارنه يتم من خلال عرض النتائج في كلا المثالين في جدول المقارنة التالي :

المؤشرات		انحدار خطي بسيط	انحدار خطي متعدد
		نوع الانحدار	
التقديرات	$\hat{a}$	12·650	3·409
	$\hat{b}_1$	− 0·950	− 0·607
	$\hat{b}_2$	−	0·05 7
معامل التحديد	R^2	0·901 (غير محتسب في حينه)	0·947
الخطأ المعياري	S_y /	0·70 (غير محتسب في حينه)	0·55

يلاحظ من الجدول اعلاه انه عند ادخال المتغير X_2 في النموذج كمتغير مؤثر في Y حصل مايلي :

أ ـ حصول تغير في قيم $\hat{a}$, $\hat{b}_1$ وعلى نحو ملحوظ بسبب تأثير X_2 على Y من خلال المعامل $\hat{b}_2$ مع بقاء اشارة $\hat{b}_1$ سالبة .

ب ـ ارتفاع في قيمة R^2 من 90.1% الى 94.7% نتيجة لاضافة مساهمة X_2 في توضيح التذبذب في Y .

جـ ـ انخفاض في قيمة الخطأ المعياري نتيجة لتقليص دور المتغير العشوائي U بسبب سحب تأثير المتغير X_2 منه من خلال تضمين X_2 في النموذج .

من خلال هذه المقارنة البسيطة يمكن القول انه كلما ازداد عدد المتغيرات المستقلة المعتمدة في النموذج التي يعتقد انها تمتلك تاثيراً على المتغير التابع فان ذلك يزيد من قيمة R^2 من جهة ويقلل من الخطأ المعياري من جهة أخرى .