

الارتباط و الانحدار و السلاسل الزمنية

الارتباط

يستهدف تحليل الارتباط معرفة إذا كانت هناك علاقة بين متغيرين أو مجموعة متغيرات مستقلة X_i و المتغير التابع، وهناك مقياسين لتحديد درجة العلاقة أو الارتباط بين المتغيرات وهما: معامل التحديد ويرمز له r^2 (R^2)، أما المعامل الثاني فيسمى معامل الارتباط (r) في حالة الارتباط البسيط و R في حالة الارتباط المتعدد) وهو عبارة عن الجذر التربيعي لمعامل التحديد.

هناك عدة أنواع من الارتباط حسب عدد المتغيرات وطبيعة العلاقة بين المتغيرات و نوع المعطيات (كمية ونوعية)، فإذا كان الارتباط خطي و المعطيات كمية (رقمية) نستخدم معامل الارتباط البسيط ومعامل الارتباط المتعدد R ومعامل الارتباط الجزئي، أما عندما تكون المعطيات نوعية (غير رقمية) فنستخدم معامل الارتباط الرتبي، r_s ومعامل الاقتران r_a ومعامل التوافق r_c إذا لم تكن المعطيات قابلة للترتيب.

1- معامل الارتباط البسيط:

وهو يعبر عن العلاقة بين متغيرين، لا يهم أيهما يكون المتغير المستقل وأيهما المتغير التابع، فقد يكون الهدف قياس العلاقة بين متغيرين مستقلين أو معتمدين أو مستقل ومعتمد، ولكن يجب أن يكون كلا المتغيرين مستقلين ويخضعان للتوزيع الطبيعي مما يمكن من تقدير معالم المجتمع، فإذا كان التوزيع غير طبيعي فلا يمكن الاستدلال بنتائج لتقدير معالم المجتمع. ويمكن حساب معامل الارتباط البسيط الذي يسمى بمعامل الارتباط بيرسون (Pearson Moment Corrélation) كالآتي:

$$r_p = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

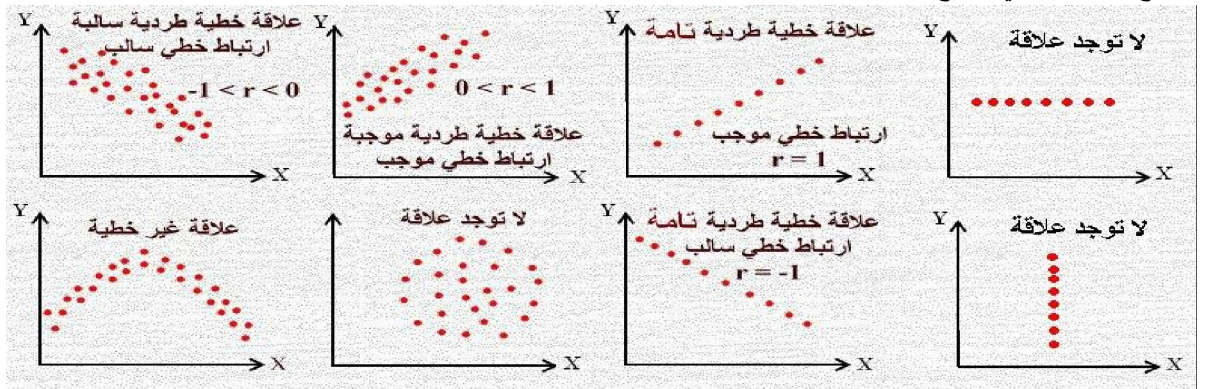
$$r_p = \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{[\sum x^2 - n\bar{x}^2][\sum y^2 - n\bar{y}^2]}}$$

$$r_p = \frac{\text{cov}(x; y)}{\delta_x \delta_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

$\text{Cov}(X, Y)$: التباين المشترك لـ X و Y

$\delta_x \cdot \delta_y$: تباين X وتباين Y

يوضح الشكل التالي انواع الارتباط



- إذا كان معامل الارتباط r يساوي الصفر: لا يوجد ارتباط
- إذا كان معامل الارتباط r يساوي الواحد: الارتباط كامل أي تام
- إذا كان معامل الارتباط r محصور بين $[0.4-0]$ ارتباط ضعيف
- إذا كان معامل الارتباط r محصور بين $[0.7-0.4]$ ارتباط متوسط
- إذا كان معامل الارتباط r محصور بين $[1-0.7]$ ارتباط قوي
- إذا كان معامل الارتباط r موجب : الارتباط طردي
- إذا كان معامل الارتباط r سالب: الارتباط عكسي

❖ اختبار المعنوية لمعامل الارتباط

لأجل اختبار معنوية معامل الارتباط البسيط بيرسون r_p ، و أن العلاقة بين المتغيرين يمثل معامل الارتباط المجتمع المسحوب من العينة نفترض أولاً أن المجتمع له توزيع طبيعي و حجم العينة صغير ($n \leq 30$) و بالتالي فإن معامل الارتباط r يكون له توزيع t (student) و بالتالي نقوم باختبار الفرضيات التالية:

$$\begin{cases} H_0 & p=0 \\ H_1 & p \neq 0 \end{cases} \quad \begin{array}{l} \text{الفرضية الصفرية} \\ \text{الفرضية البديلة} \end{array}$$

- أ- إذا كان معامل الارتباط يساوي الصفر وهو الفرض العدمي فلا يوجد علاقة أو ارتباط بين المتغيرين.
- ب- إذا كان معامل ارتباط المجتمع لا يساوي الصفر أي يوجد ارتباط بين المتغيرين نقوم في هذه الحالة بإحصائية الاختبار t كما يلي:

$$t_{cal} = r \sqrt{\frac{n-2}{1-r^2}}$$

- نقارن قيمة t الجدولية بدرجات حرية $(n-2)$ وعند مستوى المعنوية α :
- أ- إذا كانت t المحسوبة اقل من t الجدولية ، نقبل H_0 أي لا يوجد ارتباط بين المتغيرين،
- ب- إذا كان t المحسوبة اكبر من t الجدولية ، نقبل H_1 أي يوجد ارتباط بين المتغيرين.

مثال:

أجريت دراسة على معدل الدخل X ومصروفات Y لستة (6) عائلات فكانت النتائج كما في الجدول التالي:

x	10	20	30	40	50	60
y	7	21	23	34	36	53

المطلوب: إيجاد معامل الارتباط بيرسون r_p واختبار معنويته عند مستوى المعنوية $\alpha = 0.05$ ($t_{0.05}^4 = 2.132$)

الحل:

1- حساب معامل الارتباط بيرسون:

$$r_p = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

نحسب باستخدام الصيغة التالية:

لتطبيق القانون يجب إيجاد كل من $\sum xy$ ، $\sum y^2$ ، $\sum x^2$

X	Y	XY	X ²	Y ²
10	7	70	100	49
20	21	204	400	441
30	23	690	900	529
40	34	1360	1600	1156
50	36	1800	2500	1296
60	53	3180	3600	2809
210	174	7520	9100	6280

بتطبيق الصيغة نجد:

$$r_p = \frac{6(7520) - (210)(174)}{\sqrt{[6(9100) - (210)^2][6(6280) - (174)^2]}} = 0.975$$

أي هناك ارتباط طردي و قوي بين X و y

2- اختبار المعنوية:

لاختبار المعنوية نضع الفرضية التالية:

$$\begin{cases} H_0 & p=0 \\ H_1 & p \neq 0 \end{cases}$$

الفرضية الصفرية
الفرضية البديلة

$$r^2 = 0.95 \quad r = 0.975 \quad n = 6 \quad \text{لدينا:}$$

$$t_{cal} = r \sqrt{\frac{n-2}{1-r^2}} = 0.975 \sqrt{\frac{6-2}{1-0.975^2}} = 8.721$$

بالمقارنة مع قيمة t الجدولية المستخرجة من الجدول والمقدمة سابقا ($t_{0.05}^4 = 2.132$) نجد أن قيمة t المحتسبة اكبر من قيمة t الجدولية ومنه نرفض H_0 ونقبل H_1 ، ونستدل أن هناك علاقة معنوية بين المتغيرين.

2- معامل الارتباط المتعدد:

يهدف معامل الارتباط المتعدد إلى قياس العلاقة بين أكثر من متغيرين ($X_1, X_2, X_3, \dots$) ، أما اختبار معنويته فيخضع لتوزيع فيشر F.

3- معامل الارتباط الجزئي:

إن الفرق بين معامل الارتباط البسيط ومعامل الارتباط الجزئي هو أن الأول يقيس العلاقة بين متغيرين ضمن تأثيرات المتغيرات الأخرى، في حين يقيس الثاني العلاقة بين متغيرين بعد استبعاد تأثير متغير أو متغيرات أخرى بإبقائها ثابتة.

4- معامل الارتباط الرتبي:

يسمى أيضا بمعامل الارتباط لسبيرمان، ويستخدم مع المعطيات الرقمية كونه سريع الاحتساب ولكنه اقل دقة و اعتمادية من معامل لارتباط البسيط ، كذلك يستخدم مع المعطيات الكمية (غير الرقمية) القابلة للترتيب التصاعدي أو التنازلي مثل (جيد ، جيد جدا ، متوسط....) ويرمز له بالرمز r_s وهو ضمن الإحصاءات غير المعلمية ، قيمته موجبة أقل من أو يساوي الصفر (0) و تحسب قيمته بالصيغة التالية:

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

d_i الفرق بين رتبة المتغير الأول X ورتبة المتغير الثاني Y

$d_i = \text{Rank } X - \text{Rank } Y$

• ملاحظة :

- إذا كانت هناك متغيرات متكررة، لإيجاد ترتيبها نقوم بإيجاد الوسط الحسابي لترتيبها (تم التطرق إليها في التطبيق)
- لاختبار المعنوية يخضع لتوزيع t ما في معامل الارتباط البسيط (بيرسون)

مثال:

تقدم سبعة طلاب لامتحان المرحلة الثانوية وكانت معدلات نتائجهم حسب الصف و المدرسة كالتالي:

X	73	80	70	66	89	71	65
Y	69	78	66	70	92	64	55

المطلوب: حساب معامل الارتباط لسبيرمان.

الحل:

نكون جدول نوضع فيه رتب كل من X و Y ثم نحسب الفرق d_i ومربع الفرق كما في الجدول :

x	y	$\text{Rank } X$	$\text{Rank } Y$	d_i	d_i^2
73	69	3	4	-1	1
70	78	2	2	0	0
80	66	5	5	0	0
66	70	6	3	3	9
89	92	1	1	0	0
71	64	4	6	-2	4
65	55	7	7	0	0
					14

$$\begin{aligned}
 r_s &= 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \\
 &= 1 - \frac{6(14)}{7(49 - 1)} \\
 &= 0.75
 \end{aligned}$$

أي هناك ارتباط قوي و طردي بين X و Y .

5- معامل الاقتران فاي:

يستخدم معامل الاقتران " فاي " لقياس العلاقة بين المتغيرين الاسميين كل منهما ثنائي التقييم كالنوع (ذكر/أنثى) و الإصابة بالمرض (مصاب غير مصاب) و التدخين (مدخن/غير مدخن) بحيث يكون الشكل العام كما يلي:

	X ₁	X ₂	
Y ₁	a	b	a+b
Y ₂	c	d	C+d
	a+c	b+d	

معامل فاي للاقتران يعطى بالصيغة التالية:

$$r_{\phi} = \frac{a.d - b.c}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

مثال:

أوجد قيمة معامل الاقتران بين النوع (ذكر / أنثى) و الإصابة بمرض الاكتئاب (مصاب/ غير مصاب)، إذا كانت لديك المعلومات التالية :

	مصاب	غير مصاب	
ذكر	12	8	20
أنثى	4	6	10
	16	14	30

الحل:

$$a=12 \quad b=8 \quad C=4 \quad d=6$$

$$r_{\phi} = \frac{12 \times 6 - 8 \times 4}{\sqrt{20 \times 10 \times 16 \times 14}} = 0.19$$

هناك ارتباط طردي ضعيف بين x و y

الانحدار

- إن الانحدار هو أسلوب يبحث في العلاقة بين المتغيرات من خلال بناء معادلة تستخدم للتقدير و التنبؤ بقيمة المتغير التابع y بدلالة متغير آخر أو متغيرات مستقلة x و يمكن إجمال أهم أهداف تحليل الانحدار فيما يلي:
- أ- تحديد العلاقة بين المتغير التابع y والمتغير المستقل x أو أكثر،
 - ب- التنبؤ ب بدلالة قيم متغير مستقل باستخدام العلاقة التقديرية.
- وهناك عدة أنواع من الانحدار أهمها:
- الانحدار الخطي البسيط Simple Linear Regression تشير تسمية هذا المعامل "بسيط" إلى أنه يتضمن متغير تابع y يعتمد على متغير واحد مستقل x وكلمة خطي تشير إلى أن العلاقة بين المتغيرين y و x هي علاقة خطية.
 - الانحدار المتعدد Multiple Linear Regression هذا النوع من الانحدار يتضمن اعتماد المتغير y على أكثر من متغير مستقل مثل x_1 و x_2 ... الخ.
 - الإنحدار غير الخطي Non-Linear Regression إذا كانت العلاقة بين المتغير y والمتغيرات المستقلة غير خطية مثل علاقة أسية أو لوغاريتمية أو تربيعية ... الخ. وهناك أنواع أخرى مثل الانحدار الهرمي Hierarchical Regression والانحدار التدريجي Stepwise Regression وغيرها.

1- الانحدار الخطي البسيط:

هناك العديد من نماذج الانحدار قد يصل عددها إلى ما يقرب من مئات النماذج ولكن النموذج الأكثر أهمية والأكثر شيوعاً في الاستعمال هو نموذج الانحدار الخطي بسيط، في هذا النموذج يوجد لدينا المتغير التابع y المعروف أيضاً باسم متغير الاستجابة والمتغير المستقل x المعروف أيضاً باسم المتغير المتنبئ فبعد تمثيل الثنائيات المرتبة نحصل على شكل الانتشار فإذا اظهر شكل الانتشار وجود علاقة خطية بين المتغيرين نستخدم المعادلة التقديرية على النحو التالي:

$$y = a + bx + e$$

إذ أن:

y : المتغير التابع.

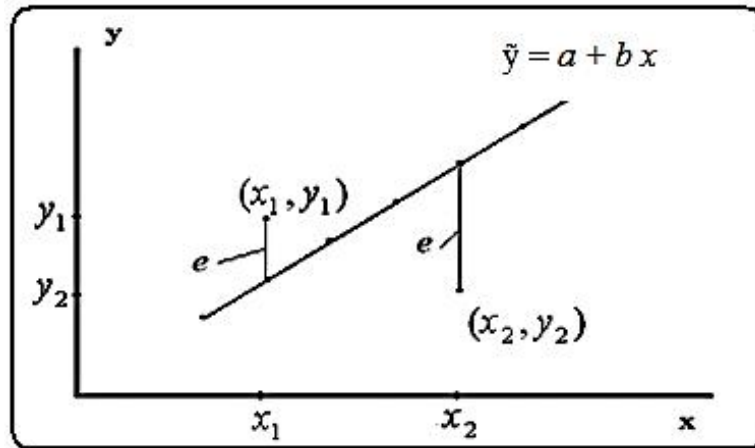
x : المتغير المستقل.

a : ثابت الانحدار وهو الجزء الذي يقطع المحور العمودي y الذي يعكس قيمة المتغير التابع y في حالة $(x = 0)$.

b : معامل الانحدار (الميل) وهو مقدار التغير في y إذا تغيرت x وحدة واحدة، ويساوي منحدر الخط المستقيم $(a + b x)$.

e : الخطأ العشوائي الذي يشير إلى الفرق بين القيمة الفعلية للمتغير التابع y والقيمة المقدرة التي يرمز لها $\tilde{y} = a + b x$,

وهذا يعني أن الخطأ العشوائي يساوي $e = y - (a + b x)$ ، ويمكن توضيح هذا الخطأ في الشكل البياني أدناه:



3- تقدير معاملات النموذج (طريقة المربعات الصغرى)

هناك عدة طرق لتقدير أو حساب نموذج الانحدار الخطي البسيط وكل الطرق تعتمد على حساب قيم معاملات الانحدار (a و b)، وتعد طريقة المربعات الصغرى Least Squares Méthode من أفضل الطرق لأنها تجعل مجموع مربعات الأخطاء العشوائية أقل ما يمكن، ولحساب القيمة المقدرة لمعامل الانحدار البسيط للمتغير التابع y بدلالة المتغير المستقل x تطبق المعادلة الآتية:

$$\hat{y} = a + b x + e$$

ولحساب قيمة a و b كما يلي:

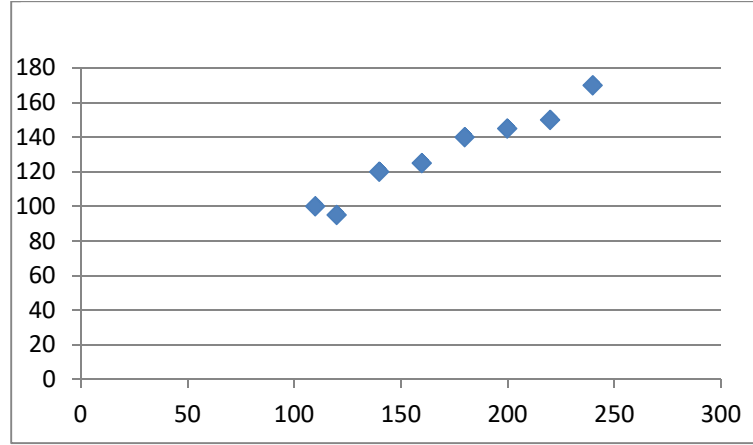
$$\begin{cases} b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} \\ a = \bar{y} - b \bar{x} \end{cases}$$

أو نستخدم قانون التباين المشترك كما يلي:

$$\begin{cases} b = \frac{\text{cov}(x,y)}{v(x)} \\ a = \bar{y} - b \bar{x} \end{cases}$$

مثال: حل التمرين الرابع من السلسلة

1- التمثيل البياني:



من خلال التمثيل البياني نستنتج أن النموذج المناسب للعلاقة بين X و Y أي بين الدخل والإنفاق علاقة طردية، حيث أن مجموعة النقاط تقع على شكل خط مستقيم، إذ يمكن اقتراح نموذج الانحدار الخطي البسيط لتمثيل العلاقة بينهما و الذي يكتب من الشكل:

$$\hat{y} = a + b x + e$$

2- تقدير معلمات النموذج:

لتقدير المعادلة Y نحسب قيمة a و b كما يلي:

$$\begin{cases} b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} \\ a = \bar{y} - b \bar{x} \end{cases}$$

نشكل الجدول:

	X : الدخل	Y : الإنفاق	XY	X ²	Y ²
1994	110	100	11000	12100	10000
1995	120	95	11400	14400	9025
1996	140	120	16800	19600	14400
1997	160	125	20000	25600	15625
1998	180	140	25200	32400	16800
1999	200	145	29000	40000	21025
2000	220	150	33000	48400	22500
2001	240	170	40800	57600	28900
	1370	1045	187200	250100	141075

✓ لتطبيق القانون نبحث أولاً على القيم XY و X^2 كما في الجدول

$$b = \frac{8(187200) - (1370)(1045)}{8(250100) - (1370)^2}$$

$$b = 0.53$$

$$\bar{x} = \frac{\sum x}{n} = \frac{1370}{8} = 171.25$$

$$\bar{y} = \frac{\sum y}{n} = \frac{1045}{8} = 130.625$$

$$a = 130.625 - 0.53(171.25)$$

$$a = 39.8 \approx 40$$

ومنه المعادلة التقديرية هي: $y = 0.53x + 39.8$

3- حساب تأثير الدخل على الاستهلاك:

لحساب تأثير الدخل على الاستهلاك يجب حساب معامل التحديد (مربع معامل الارتباط r^2)

$$r_p = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

$$r_p = \frac{8(187200) - (1370)(1045)}{\sqrt{(123900)(36575)}}$$

$$r_p = 0.979$$

ومنه معامل التحديد $r^2 = 0.96$

أي 96% من التغير في y سببه التغير في x و الباقي يعود لأسباب أخرى نجهلها

السلاسل الزمنية

السلاسل الزمنية تتكون من أربعة عناصر أساسية وهي:

- معادلة الاتجاه العام
- التغيرات الدورية
- التغيرات الموسمية
- التغيرات الفجائية