وزارة التعليم العالي والبحث العلمي

جامعة تكريت

كلية الإدارة والاقتصاد

قسم ادارة الاعمال

المرحلة الاولى

محاضرات في مادة الاحصاء

مدرس المادة

أ.م.د هادي مران احمد

1446 هـ **2024 مـ**

الفصل الخامس

الارتباط والانحدار

الارتباط الخطي :

إن مفهوم الارتباط الخطي يقترن بحالة وجود متغيرين أو أكثر يرتبطان مع بعضهما بعلاقات خطية معينة ، على سبيل المثال العلاقة بين طول الشخص ووزنه ، فإذا كان التغيير في إحدى المتغيرات يؤثر في تغيير متغير آخر أو مجموعة متغيرات عندئذ يقال أن هذه المتغيرات مرتبطة فيما بينها ، وإذا كان المتغيرين يتغيران بنفس الاتجاه أي زيادة (نقصان) في أحدها تؤدي إلى زيادة (نقصان) في الآخر عندئذ يقال أن الارتباط فيما بينهما هو موجب ، أما إذا كان المتغيرين المرتبطين أو مجموعة المتغيرات يتغيران باتجاه معاكس أي زيادة (نقصان) في أحدهما يؤدي إلى نقصان (زيادة) في الآخر عندئذ يقال الارتباط هو سالب.

الارتباط الخطي البسيط : يعرف بأنه درجة أو القيمة العددية للعلاقة بين متغيرين وفيما يلي المفاهيم الأساسية التي توضح فكرة الارتباط البسيط.

اولاً : معامل الارتباط الخطي (بيرسون)

حساب معامل الارتباط البسيط للبيانات: وعليه فأن معامل الارتباط يكون مساوي للقانون التالي :

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

مثال 1 : البيانات التالية تمثل الكمية المعروضة من سلعة معينة وسعر الوحدة الواحدة من هذه السلعة.

المطلوب : حساب معامل الارتباط البسيط بين الكمية المعروضة والسعر.

السعر x	الكمية المعروضة y	Xy	x^2	y^2
2	3	6	4	9
2	5	10	4	25
5	7	35	25	49

4	8	32	16	64
5	9	45	25	81
6	11	66	36	121
3	6	18	9	36
5	8	40	25	64
4	6	24	16	36
36	63	276	160	485

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

$$= r_p = \frac{9\,(276)-(36)(63)}{\sqrt{(9(160)-36^2\,)\,(9\,(485)-63^2\,)}}$$

$$= \frac{2484-2268}{\sqrt{(1440-1296\,)\,(4365-3969\,)}}$$

$$= \frac{216}{\sqrt{(144\,)\,(396\,)}}$$

$$= \frac{216}{\sqrt{(144\,)\,(396\,)}} = \frac{216}{\sqrt{57024}}$$

$$= \frac{216}{238.79} = 0.90$$

الارتباط طردي قوي

وبالإمكان تقسيم الارتباط الخطي إلى :

أ- الارتباط الخطي البسيط : وهو الارتباط بين الظاهرتين كما ذكرنا.

ب- الارتباط الجزئي : وهو الارتباط بين ظاهرتين بثبوت الظاهرة الثالثة.

ج- الارتباط المتعدد : وهو الارتباط بين ثلاث ظواهر.

من المعلوم أن مقياس الارتباط يختلف حسب اختلاف الظواهر المقاسة والتي يراد إيجاد العلاقة بينهما فإذا كانت لدينا الحالات التالية :

1- إذا كانت الظاهرتين مقاسة قياساً كمياً أي بالإمكان التعبير عنها بالأرقام نستخدم صيغة بيرسون للارتباط كما موضح سابقاً.

2- إذا كانت إحدى الظواهر المقاسة لا يمكن التعبير عنها كمياً مثل الحالة العلمية مع الدخل سوف نستخدم صيغة (سبيرمان أو كندال).

ثانياً : معامل ارتباط الرتب لسبيرمان Spearman Rank correlation coff. :

أفرض أن y , x متغيرين من النوع الوصفي وأفرض أن البيانات المستحصل عليها على أساس عينة عشوائية من المفردات قوامها n هي بهيئة صفات غير قابلة للقياس الكمي ويسمى الارتباط لهذه الحالة كما تكرر بارتباط الرتب والصيغة المستخدمة لإيجاد معامل الارتباط هي :

معامل ارتباط الرتب

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

حيث أن : r تمثل معامل الارتباط (ارتباط الرتب).

n : عدد القيم لكلا الظاهرتين.

$\sum_{i=1}^{n} d_i^2$: عبارة عن مجموع مربعات الفروق بين رتب الظاهرة الأولى ورتب الظاهرة الثانية المناظرة لها.

أما الخطوات الواجب اتباعها لتطبيق هذه الصيغة هي كما يلي :

1- نعطي قيم من الأعداد الطبيعية على قيم الظاهرتين مباشرة بترتيب تصاعدي.

2- نعطي القيم المتكررة معدل من الرتب التي تناظرها.

3- نرتب قيم الظاهرة الأولى بتسلسل الأعداد الطبيعية وتذكر الرتب التي تناظرها من رتب الظاهرة الثانية.

4- تحسب الفروق بين الرتب والرتب المناظرة فنحصل على $\sum di$ ثم نربع هذه الفروق ثم نجمعها والخطوة الأخيرة نطبق الصيغة المذكورة أعلاه.

ملاحظة : يمكن استخدام صيغة (سبيرمان) لإيجاد معامل ارتباط الرتب بين ظاهرتين غير مقاسة كمياً ، وذلك بذكر تسلسل الظاهرتين تصاعدياً وإعطاءهما رتب من الاعداد الطبيعية ثم نكمل الحل بإتباع الخطوات المذكورة أعلاه.

مثال 2 :

لديك البيانات التالية والمطلوب معرفة العلاقة بين الظاهرتين (y , x) وتفسير النتيجة ، علماً أن البيانات تمثل الحالة الانتاجية لأحد المصانع وكمية الإنتاج لإحدى السلع في هذا المصنع.

x_i	y_i	الرتب x_i والرتب المناظرة y_i	d_i	d_i^2
$^{3}76$	متوسط^{2}	3 2	1	1
$^{4}82$	جيد^{3}	4 3	1	1
$^{1}63$	ضعيف^{1}	1 1	0	0
$^{2}70$	جيد جداً^{4}	2 4	2	4
				$6 \rightarrow \sum_{i=1}^{n} d_i^2$

$$r_{xy} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6\,(6)}{4\,(16-1)} = 1 - \frac{36}{(4)(15)} = 1 - \frac{36}{60} = 1 - 0.60 = 0.40$$

مثال 3 : أجاب عشرة موظفين عن حالتهم العلمية ومقدار دخلهم الشهري بالدينار فكانت إجاباتهم كما في الجدول أدناه ، المطلوب إيجاد معامل ارتباط الرتب باستخدام صيغة (سبيرمان) بين الدخل الشهري والحالة العلمية وتفسير النتائج.

الحالة العلمية xi	الدخل الشهري y_i	الرتب x_i والرتب المناظرة إلى y_i	d^i	d^{i2}
يقرأ ويكتب ^{4}	$^{7}43$	4 − 7	−3	9
متوسطة ^{7}	$^{6}39$	7 − 6	1	1
أمي$^{1.5}$	$^{2}11$	1.5 − 2	−0.5	0.25
يقرأ ويكتب ^{4}	$^{3}16$	4 − 3	1	1
عليا $^{9.5}$	$^{1}10$	9.5 − 1	8.5	2.25
عليا $^{9.5}$	$^{9}78$	9.5 − 9	0.5	0.25
متوسطة ^{7}	$^{10}87$	7 − 10	−3	9
يقرأ ويكتب ^{4}	$^{5}32$	4 − 5	−1	1
أمي$^{1.5}$	$^{4}19$	1.5 − 4	−2.5	6.25
متوسطة ^{7}	$^{8}50$	7 − 8	−1	1
المجموع				101

$$r = 1 - \frac{6\sum_{i=1}^{n} di^2}{n\,(n^2-1)} = 1 - \frac{6\,(101)}{10\,(10-1)} = 1 - \frac{606}{990} = 0.39$$

العلاقة ضعيفة بين الظاهرتين.

الانحدار The Regression:

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

يعرف الانحدار أو تحليل الانحدار بشكل عام بأنه مقياس رياضي لمتوسط العلاقة بين متغيرين أو أكثر بدلالة وحدات قياس المتغيرات المعتمدة في العلاقة وغالباً ما تسمى العلاقات من هذا النوع بنماذج الانحدار.

ويختص الانحدار الخطي بتحليل العلاقة بين متغيرين أو أكثر يظهر كل منهما ب (أس مساو للواحد) في العلاقة المعتمدة للتحليل وتكون المعادلة بالشكل التالي :

$$y_i = a + bx_i + e_i \rightarrow \text{نموذج معادلة الانحدار}$$

حيث أن b , a معالم النموذج ، y يمثل المتغير المعتمد للنموذج ، علماً أن كل من y , x متغيرات من الدرجة الأولى (أس واحد) ولتقدير معادلة نموذج الانحدار أعلاه باستخدام أحد طرق التقدير نحصل على :

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

حيث أن $\hat{a}, \hat{b}$ مقدرات المعالم b , a وتلفظ (a hat) ، (b hat) أما هذا الخطأ العشوائي e_i فأنه يكون مساوي إلى :

متغير الخطأ العشوائي $e_i = y_i - \hat{y}_i$

بتربيع المعادلة أعلاه ينتج :

$$e_i^2 = (y_i - \hat{y}_i)^2$$

بإدخال المجموع نحصل على :

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

حيث ان:

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

$$\therefore \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{a} - \hat{b}x_i)^2 \qquad ,,,,,,,,,,(1)$$

الهدف هنا إيجاد مقدرات $\hat{a},\hat{b}$ التي تجعل $\sum_{i=1}^{n} e_i^2$ أقل ما يمكن ، وعليه فأن :

$\bar{y} = \hat{a} - \hat{b}\bar{x}$ ويسمى انحدار y / x

وبعد إجراء عدة اشتقاقات على المعادلة (1) نحصل على تقديرات $\hat{a},\hat{b}$ مساوية أي :

$$\hat{b} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

$$= \bar{y} - \hat{b}\bar{x}\hat{a}$$

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

مثال 7 :

البيانات التالية تمثل الكمية المطلوبة (y) من سلعة معينة وسعر الوحدة الواحدة منها x .

المطلوب:

1- تقدير معادلة انحدار الكمية المعروضة على السعر ؟

2- تقدير الكمية المعروضة عند سعر 20 ؟

X	Y	X y	x^2
7	10	70	49
10	15	150	100
4	5	20	16
3	4	12	9
2	3	6	4
9	13	117	81
35	50	375	259

$\bar{X} = \frac{35}{6} = 5.83$

$\bar{Y} = \frac{50}{6} = 8.33$

$\hat{b} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{6(375)-(35)(50)}{6(259)-35^2} = \frac{2250-1750}{1554-1225} = \frac{500}{329} = 1.51$

$= \bar{y} - \hat{b}\bar{x} \quad \hat{a}$

$= 8.33 - 1.51\ (5.83)$

$= 8.33 - 8.80 = -0.47$

معادلة الانحدار المقدرة هي : $= \hat{a} + \hat{b}x \ \hat{y}$

وتقدير الكمية المعروضة عند سعر (20) يكون عن طريق تعويض معادلة الانحدار المقدرة ($x = 20$)

$= -0.47 + 1.51\ (20)\ \hat{y}$

$= -0.47 + 30.2 = 27.73$