المحاضرة السابعة

الارتباط الخطي البسيط

الدرس الخامس

الارتباط والانحدار الخطي البسيط

- **مقدمة.**

بعد استعراض العديد من مقاييس وصف البيانات والتي تقوم بوصف البيانات الخاصة بمتغير واحد مثل إنتاج مصنع معين آو كمية استهلاك سلعة غذائية معينة....الخ. ألان وفي هذا الدرس سوف نتطرق إلي دراسة متغيرين آو أكثر في أن واحد وذلك باستخدام أهم تقنيات علم الإحصاء في تحليل البيانات وهو تحليل الارتباط و الانحدار الخطي البسيط، هذه التقنية في التحليل الإحصائي تقوم علي دراسة العلاقة بين متغيرين بواسطة تحليل الارتباط أي دراسة ارتباط المتغير **x** والمتغير **y** وهل توجد علاقة فعليا بينهما ، كذلك يمكننا دراسة اثر المتغير **x** علي المتغير **y** وذلك باستخدام تحليل الانحدار علي سبيل المثال يمكننا دراسة اثر متغير الدخل الشهري علي متغير الإنفاق. ودراسة متغير وزن الجسم ومعدل الوجبات الغذائية ....الخ.

أولا- الارتباط الخطي البسيط Simple Correlation

كما ذكر سابقا فان تحليل الارتباط الهدف منه هو دراسة و تحديد نوع وقوة العلاقة التي تربط متغيرين، لذلك وفي حالة افتراض أن العلاقة بين المتغيرين تأخذ الشكل الخطي فيجب حساب هذا الارتباط في حالة البيانات الكمية وفي حالة البيانات الوصفية المقاصة بواسطة مقياس ترتيبي، و يرمز للارتباط الخطي البسيط في حالة دراسة مفردات من المجتمع بالرمز r وكذلك يرمز له أيضا في حالة العينة، وبالتالي فان الغرض من هذا النوع من التحليل ينحصر في محورين أساسين وهما كالتي.

- **نوع العلاقة:ـــ** وتأخذ ثلاث أنواع حسب إشارة معامل الارتباط كما يلي:

1- إذا كانت إشارة معامل الارتباط سالبة ($r < 0$) توجد علاقة عكسية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه انخفاض في المتغير الثاني، والعكس.

2- إذا كانت إشارة معامل الارتباط موجبة ($r > 0$)توجد علاقة طردية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه زيادة في المتغير الثاني، والعكس .

3- إذا كان معامل الارتباط قيمته صفرا ($r = 0$) دل ذلك على انعدام العلاقة بين المتغيرين.

- **قوة العلاقة:ـــ** ويمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن (±1)، حيث أن قيمة معامل الارتباط تقع في المدى($-1 < r < 1$)، وقد صنف بعض الإحصائيين درجات لقوة العلاقة يمكن تمثيلها على الشكل التالي:

درجات قوة معامل الارتباط

- معامل الارتباط الخطي البسيط لبيرسون Pearson

أن معامل الارتباط الخطي البسيط لبيرسون يستخدم للقياس وتحليل البيانات الكمية والعددية، بحيث أذا كانت لدينا بيانات للمتغيرين من عينة هما (x . y) فانه يمكن قياس الارتباط بينهما بواسطة طريقة بيرسون، مثال علي ذلك يمكن قياس الارتباط بين متغير الوزن والمتغير الطول ، وكذلك قياس الارتباط بين متغير التسميد ومتغير الإنتاج ألفلاحي ، وأيضا قياس الارتباط بين متغير اليد العاملة الماهرة ومتغير سرعة الانجاز.....الخ. وبالتالي لحساب معامل الارتباط بطريقة بيرسون نستخدم الطريقة المعدلة التالية.

$$r = \frac{S_{xy}}{S_x\, S_y} = \frac{\dfrac{\sum (x-\bar{x})(y-\bar{y})}{(n-1)}}{\sqrt{\dfrac{\sum (x-\bar{x})^2}{(n-1)}}\sqrt{\dfrac{\sum (y-\bar{y})^2}{(n-1)}}}$$

حيث آن

$S_{xy} = \sum (x-\bar{x})(y-\bar{y})/(n-1)$: هو التغاير بين $(y\ .\ x)$،

$S_x = \sqrt{\sum (x-\bar{x})^2/(n-1)}$: هو الانحراف المعياري لقيم (x)،

$S_y = \sqrt{\sum (y-\bar{y})^2/(n-1)}$: هو الانحراف المعياري لقيم (y).

هذه العلاقة يمكن تلخصيها إلي العلاقة الآتية.

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2}\sqrt{\sum (y-\bar{y})^2}}$$

مثال

فيما يلي المساحة المزرعة بالأعلاف الخضراء بالألف هكتار، وإجمالي إنتاج اللحوم بالألف طن، خلال الفترة من 1995حتى عام 2002.

السنة	1995	1996	1997	1998	1999	2000	2001	2002
المساحة	305	313	297	289	233	214	240	217
الكمية	592	603	662	607	635	699	719	747

والمطلوب: حساب معامل الارتباط بين المساحة والكمية، وما هو مدلوله ؟

الحل

بفرض أن (x) هي المساحة المزرعة، (y) هي الكمية، ولحساب معامل الارتباط بين $(x \ , \ y)$ يتم تطبيق المعادلة (6-2)، وذلك على النحو التالي:

- حساب الوسط الحسابي لكل من المساحة، والكمية $(\bar{x} \ , \ \bar{y})$.

$$\bar{x} = \frac{\sum x}{n} = \frac{2108}{8} = 263.5 \ , \ \bar{y} = \frac{\sum y}{n} = \frac{5264}{8} = 658$$

- حساب المجاميع

x	y	$x-\bar{x}$	$(x-\bar{x})^2$	$y-\bar{y}$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
305	592	41.5	1722.25	-66	4356	-2739
313	603	49.5	2450.25	-55	3025	-2722.5
297	662	33.5	1122.25	4	16	134
289	607	25.5	650.25	-51	2601	-1300.5
233	635	-30.5	930.25	-23	529	701.5
214	699	-49.5	2450.25	41	1681	-2029.5
240	719	-23.5	552.25	61	3721	-1433.5
217	747	-46.5	2162.25	89	7921	-4138.5
2108	5264	0	12040	0	23850	-13528

$$\sum(x-\bar{x})^2 = 12040\ ,\ \sum(y-\bar{y})^2 = 23850\ ,$$
$$\sum(x-\bar{x})(y-\bar{y}) = -13528$$

إذا معامل الارتباط قيمته هي:

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}} = \frac{-13528}{\sqrt{12040}\sqrt{23850}}$$
$$= \frac{-13528}{(109.727)(154.434)} = \frac{-13528}{16945.619} = -0.798$$

- يوجد ارتباط عكسي قوي بين المساحة المزرعة، وكمية إنتاج اللحوم.

ولتبسط العلاقات الحسابية السابقة يمكننا تطبيق العلاقة الآتية، والتي تعتمد علي مجموع القيم وليس علي انحرافات القيم عن وسطها الحسابي.

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

وبالتطبيق على بيانات المثال السابق ، يتبع الآتي :

- حساب المجاميع:

x	y	xy	x^2	y^2
305	592	180560	93025	350464
313	603	188739	97969	363609
297	662	196614	88209	438244
289	607	175423	83521	368449
233	635	147955	54289	403225
214	699	149586	45796	488601
240	719	172560	57600	516961
217	747	162099	47089	558009
2108	5264	1373536	567498	3487562

المجاميع المطلوبة
$\sum x = 2108 \ , \ \sum y = 5264$ $\sum xy = 1373536$
$\sum x^2 = 567498$ $\sum y^2 = 3487562$

- حساب معامل الارتباط:

باستخدام المجاميع السابقة، وبالتطبيق على المعادلة (3-6) أعلاه، نجد أن معامل الارتباط قيمته هي:

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

$$= \frac{1373536 - \frac{(2108)(5264)}{8}}{\sqrt{\left(567498 - \frac{(2108)^2}{8}\right)\left(3487562 - \frac{(5264)^2}{8}\right)}}$$

$$= \frac{-13528}{\sqrt{(12040)(23850)}} = \frac{-13528}{16945.619} = -0.798$$

أما في حالة المتغيرات الوصفية فأننا نستخدم طريقة سيبرمان لدراسة لتحليل الارتباط بين المتغيرات.

- **معامل ارتباط الرتب اسبيرمان Spearman .**

لدراسة عينة ذات متغيرات وصفية مثل دراسة الارتباط لمتغير مدي الرضاء عن درجات الحرارة آو دراسة العلاقة بين نوع التقديرات الشعور عن الراحة الحرارية آو تقيم مستوي الطلبة في الجامعة فأننا نستخدم طريقة اسبيرمان لحساب الارتباط بين المتغيرات والذي يعتمد علي رتب مستويات القيم للمتغيرين كبديل للقيم الأصلية ويعبر عم طريقة اسبيرمان بعلاقة الرياضية الآتية.

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

حيث أن d هي الفرق بين رتب مستويات المتغير الأول x، ورتب مستويات المتغير الثاني y،
أي أن : $d = R_x - R_y$.

مثال علي ذلك

فيما يلي تقديرات 10 طلاب في مادتي الإحصاء، والاقتصاد:

تقديرات إحصاء	أ	جـ+	د	د+	ب+	جـ+	أ+	ب	ب+	ب+
تقديرات اقتصاد	أ+	د	جـ	جـ	أ	ب	ب+	ب	جـ	ب

والمطلوب:

1- احسب معامل الارتباط بين تقديرات الطلبة في المقررين.

2- وما هو مدلوله ؟

الحل

1- بفرض أن x هي تقديرات الإحصاء، y هي تقديرات الاقتصاد، يمكن حساب معامل الارتبـــاط بينهما باستخدام المعادلة (6-4)، وذلك بإتباع الآتي:

الرتب	1	2	3	4	5	6	7	8	9	10
تقديرات إحصاء	أ+	أ	ب+	ب+	ب+	ب	جـ+	جـ+	د+	د
رتب x	1	2	(3+4+5)/3=4			6	(7+8)/2=7.5		9	10
تقديرات اقتصاد	أ+	أ	ب+	ب	ب	ب	جـ	جـ	جـ	د
رتب y	1	2	3	(4+5+6)/3=5			(7+8+9)/3=8			10

- إذا يمكن حساب المجموع: $\sum d^2$ كما يلي:

x	y	رتب x	رتب y	d	d^2
أ	أ+	2	1	1	1
جـ+	د	7.5	10	-2.5	6.25
د	جـ	10	8	2	4
د+	جـ	9	8	1	1
ب+	أ	4	2	2	1
جـ+	ب	7.5	5	2.5	6.25
أ+	ب+	1	3	-2	4
ب	ب	6	5	1	1
ب+	جـ	4	8	-4	16
ب+	ب	4	5	-1	1
					44.5

$$\sum d^2 = 44.5$$

- معامل الارتباط هو:

$$r = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$$= 1 - \frac{6(44.5)}{10(10^2-1)} = 1 - \frac{267}{990}$$

$$= 1 - 0.2697 = 0.7303$$

2- مدلول معامل الارتباط :

بما أن $r = 0.703$ ، ويدل ذلك على وجود ارتباط طردي قوي بين تقديرات الطالـــب في مادة الإحصاء ، ومادة الاقتصاد .