الارتباط والانحدار Correlation and Regression

الارتباط

الارتباط هو مقياس يستخدم لقياس درجة العلاقة بين متغيرين وبيان مدى تأثير احدهما على الآخر، على سبيل المثال تركيز الأوكسجين في الماء يعتمد بشكل كبير على درجة حرارة الماء، ارتفاع درجة الحرارة يؤدي إلى انخفاض تركيز الأوكسجين والعكس صحيح، في هذا المثال لدينا متغيرين الأول هو تركيز الأوكسجين في الماء ويدعى في هذه الحالة متغير معتمد Dependent variable لأن تركيز الأوكسجين في الماء يعتمد على درجة حرارة الماء، بينما درجة حرارة الماء يسمى متغير مستقل Independent variable وذلك لأن درجة حرارة الماء لا تتأثر بتغير تركيز الأوكسجين فيه. من الأمثلة الأخرى علاقة الطول بالوزن في الأسماك، زيادة الطول تؤدي إلى زيادة الوزن بينما ليس بالضرورة أن تكون زيادة الوزن مصحوبة بزيادة في الطول، لذلك تعتبر الزيادة في الطول متغير مستقل بينما الزيادة في الوزن متغير معتمد. تقاس قوة العلاقة بين المتغيرين اعتمادا على قيمة معامل الارتباط Correlation coefficient (r) والذي تتراوح قيمته بين +1 إلى −1 إذ كلما اقتربت القيمة من الواحد كلما كانت علاقة الارتباط أقوى بينما كلما اقتربت القيمة من الصفر كلما كانت علاقة الارتباط أضعف. عندما تكون قيمة معامل الارتباط موجبة فان العلاقة بين المتغيرين تكون طردية أي إن زيادة المتغير المستقل تؤدي إلى زيادة المتغير المعتمد كما هو الحال في علاقة الطول بالوزن، بينما عندما تكون قيمة معامل الارتباط سالبة فان العلاقة بين المتغيرين تكون عكسية أي إن زيادة المتغير المستقل تؤدي إلى تناقص المتغير المعتمد كما هو الحال في علاقة درجة حرارة الماء وتركيز الأوكسجين فيه.

يجب التأكيد على انه عند دراسة الارتباط بين متغيرين يجب ان يكون هناك تفسير منطقي لاختيار المتغيرين اذ ان من الصعب مثلا تفسير الارتباط بين التدخين والدرجات الامتحانية للطلاب. لذلك يمكن القول ان معامل الارتباط هو من اكثر المقاييس الاحصائية عرضة للوقوع في تفسيرات خاطئة.

تقدير معامل الارتباط

تستخدم الصيغة التالية لقياس معامل الارتباط r

$$r = \frac{\sum X_i Y_i - \frac{(\sum X_i)(\sum Y_i)}{n}}{\sqrt{(\sum X_i^2 - \frac{(\sum X_i)^2}{n})(\sum Y_i^2 - \frac{(\sum Y_i)^2}{n})}}$$

بما ان قيمة r تتراوح بين -1 و +1 اي ان

$-1 \leq r \leq +1$

من اهم مزايا معامل الارتباط انه لا يتأثر بوحدات القياس فهو عدد مجرد خال من وحدات القياس.

مثال/ أجريت تجربة لمعرفة علاقة محتوى الكبد في اسماك الكارب الشائع من الكلايكوجين (متغير مستقل) مع قيمة الدالة الكبدية (متغير معتمد).

محتوى الكبد من الكلايكوجين (X): 16.10، 24.90، 33.01، 37.88

الدالة الكبدية (Y): 0.84، 1.22، 1.26، 1.44

هل هناك علاقة بين المتغيرين وما هي طبيعة هذه العلاقة؟

الحل/

1- ننشيء جدول نستخرج منه القيم المطلوبة في معادلة معامل الارتباط

المكرر	X_i	Y_i	X_iY_i	X_i^2	Y_i^2
1	16.10	0.84	13.52	259.21	0.71
2	24.90	1.22	30.38	620.01	1.49
3	33.01	1.26	41.59	1089.66	1.59
4	37.88	1.44	54.55	1434.89	2.07
المجموع	**111.89**	**4.76**	**140.04**	**3403.77**	**5.86**

2- نطبق المعادلة لاستخراج قيمة r

$$r = \frac{140.0418 - \frac{(111.89)(4.76)}{4}}{\sqrt{\left(3403.7745 - \frac{(111.89)^2}{4}\right)(5.8552 - \frac{(4.76)^2}{4})}}$$

$$r = \frac{140.04 - \frac{532.60}{4}}{\sqrt{\left(3403.7745 - \frac{(12519.3721)}{4}\right)(5.8552 - \frac{(22.6576)}{4})}}$$

$$r = \frac{140.04 - 133.15}{\sqrt{(3403.7745 - 3129.8430)(5.8552 - 5.6644)}}$$

$$r = \frac{6.89}{\sqrt{(273.9315)(0.1908)}}$$

$$r = \frac{6.89}{\sqrt{52.2661}}$$

$$r = \frac{6.89}{7.2295}$$

$$r = 0.95$$

اذن قيمة معامل الارتباط r يساوي 0.95 + وبما انه قريب جدا من الواحد فأذن هناك علاقة ارتباط قوية وطبيعة العلاقة طردية كون الناتج موجب بمعنى ان زيادة المتغير X تؤدي الى الى زيادة المتغير Y.

ملاحظة: الاعتماد على قيمة معامل الارتباط بحد ذاتها للاستنتاج فيما اذا كانت توجد علاقة بين المتغيرين او لا توجد علاقة بينهما قد يكون غير دقيق تماما اذ احيانا على سبيل المثال عندما تكون قيمة معامل الارتباط تساوي رقما معينا فان الاعتماد على قرب او بعد هذا الرقم عن سالب واحد او موجب واحد للحكم على وجود او عدم وجود علاقة ارتباط يمكن ان يؤدي الى استنتاج خاطيء لذلك يجب ان تجرى اختبارات لمعرفة ما اذا كانت قيمة معامل الارتباط معنوية او غير معنوية من خلال استخراج قيمة t لكن الطريقة الاسهل هي مقارنة قيمة معامل الارتباط التي حصلنا عليها مع قيم معدة

مسبقا موجودة في كتب الاحصاء فاذا كانت قيمة معامل الارتباط المستحصل عليها اكبر او تساوي القيمة المناظرة لها في الجداول حسب درجات الحرية ومستوى المعنوية المفترض للاختبار عندها تكون قيمة معامل الارتباط معنوية بينما اذا كانت اقل عندها تكون قيمة معامل الارتباط غير معنوية.

مثال/ أجريت تجربة لمعرفة علاقة درجة حرارة الماء (متغير مستقل) مع قيمة تركيز الاوكسجين الذائب في الماء (متغير معتمد) في احد الاحواض.

درجة حرارة الماء (X): 20.4، 24.8، 31.2، 34.9

تركيز الاوكسجين الذائب في الماء (Y): 8.1، 7.5، 6.9، 6.2

هل هناك علاقة بين المتغيرين وما هي طبيعة هذه العلاقة؟

الحل/

1- ننشيء جدول نستخرج منه القيم المطلوبة في معادلة معامل الارتباط

المكرر	X_i	Y_i	X_iY_i	X_i^2	Y_i^2
1	20.4	8.1	165.24	416.16	65.61
2	24.8	7.5	186.00	615.04	56.25
3	31.2	6.9	215.28	973.44	47.61
4	34.9	6.2	216.38	1218.01	38.44
المجموع	**111.30**	**28.70**	**782.90**	**3222.65**	**207.91**

2- نطبق المعادلة لاستخراج قيمة r

$$r = \frac{782.90 - \frac{(111.30)(28.70)}{4}}{\sqrt{\left(3222.65 - \frac{(111.30)^2}{4}\right)\left(207.91 - \frac{(28.70)^2}{4}\right)}}$$

$$r = \frac{782.90 - \frac{3194.31}{4}}{\sqrt{\left(3222.65 - \frac{(12387.69)}{4}\right)\left(207.91 - \frac{(823.69)}{4}\right)}}$$

$$r = \frac{782.90 \ - 798.5775}{\sqrt{(3222.65 - 3096.9225)\ (207.91 - 205.9225)}}$$

$$r = \frac{-\ 15.6775}{\sqrt{(125.7275)\ (1.9875)}}$$

$$r = \frac{-\ 15.6775}{\sqrt{249.8834}}$$

$$r = \frac{-\ 15.6775}{15.8077}$$

$$r = -0.99$$

اذن قيمة معامل الارتباط r يساوي 0.99 - وبما انه قريب جدا من الواحد فأذن هناك علاقة ارتباط قوية وطبيعة العلاقة عكسية كون الناتج سالب بمعنى ان زيادة المتغير X تؤدي الى الى نقصان المتغير Y.