الارتباط Correlation

الارتباط لغوياً يعني التلازم، ويمكن تعريفه: بأنّه تلازم صفتين تلازماً كمياً بحيث أن تغيّر الصفة الأولى يؤدي إلى تغير الصفة الثانية، ويكون التغير بالطبع بالزيادة أو النقصان. ولا يعدُّ وجود ارتباط بين المتغيرين مؤشّراً على أن أحدهما يحدث نتيجةً لتأثير المتغير الآخر، ولا يحدث إلّا في وجوده أو بسببه، حيث يمكن أن تكون هناك عوامل تؤثّر على المتغيرين معاً فتجعل التغير الذي يحدث في أحدهما كأنه ناتج من التغير الذي حدث في المتغير الثاني.

من تعريف الارتباط نلاحظ وجود عدة احتمالات:

1. زيادة الصفة أو الظاهرة الأولى تقود إلى زيادة الصفة أو الظاهرة الثانية، وبالتالي نقول أن لدينا علاقة ارتباط طردية بين الظاهرتين وأن الارتباط موجب أي أن زيادة الصفة الأولى أدّت إلى زيادة الصفة الثانية.

مثال: زيادة كمية السماد المضافة تؤدي إلى زيادة الإنتاجية وبالتالي لدينا علاقة طردية والارتباط موجب.

أو: زيادة كمية مياه الري تؤدي إلى زيادة الإنتاجية وبالتالي لدينا علاقة طردية وارتباط موجب.

2. زيادة الصفة أو الظاهرة الأولى تقود إلى نقصان الصفة أو الظاهرة الثانية، وبالتالي نقول أن لدينا علاقة ارتباط عكسية بين الظاهرتين وأن الارتباط سالب أي أن زيادة الصفة الأولى أدّت إلى نقصان الصفة الثانية.

مثال: زيادة الإصابة بالأمراض والحشرات تؤدي إلى نقصان الإنتاجية وبالتي لدينا علاقة عكسية والارتباط سالب.

3. نقصان الصفة أو الظاهرة الأولى تقود إلى نقصان الصفة أو الظاهرة الثانية، وبالتالي نقول أن لدينا علاقة ارتباط طردية بين الظاهرتين وأن الارتباط موجب أي أن نقصان الصفة الأولى أدّت إلى نقصان الصفة الثانية.

مثال: نقصان كميّة البذار تؤدي إلى نقصان الإنتاجية وبالتالي لدينا علاقة طردية والارتباط موجب.

أو: نقصان كمية الأعلاف المقدمة للحيوانات الزراعية تؤدي إلى نقصان الإنتاجية وبالتالي لدينا علاقة طردية والارتباط موجب.

4. نقصان الصفة أو الظاهرة الأولى تقود إلى زيادة الصفة أو الظاهرة الثانية، وبالتالي نقول أن لدينا علاقة ارتباط عكسية بين الظاهرتين وأن الارتباط سالب أي أن نقصان الصفة الأولى أدّت إلى زيادة الصفة الثانية.

مثال: نقصان تكاليف الإنتاج تؤدي إلى زيادة الربح وبالتالي لدينا علاقة عكسية والارتباط سالب.

كل ما تكلّمنا عنه سابقاً يدعى الارتباط البسيط Simple Correlation حيث ندرس العلاقة بين متغيرين فقط، وهو ما سنركّز عليه في دراستنا لأهميته وبساطته مقارنةً بأنواع الارتباط الأخرى.

أنواع الارتباط: تشمل ما يلي:

1. **الارتباط البسيط:** يستخدم الارتباط البسيط لدراسة العلاقة بين متغيرين احتماليين فقط.
2. **ارتباط الرتب:** يستخدم ارتباط الرتب لدراسة العلاقة بين المتغيرات الاحتمالية عندما تكون بيانات هذه العوامل مرتّبة أو يمكن ترتيبها في رتب معينة.
3. **الارتباط الجزئي:** يستخدم الارتباط الجزئي لدراسة العلاقة بين المتغيرات الاحتمالية عندما يكون لدينا عدة متغيرات ونرغب بحساب الارتباط بين متغيرين محددين علماً بأن الارتباط مع بقية المتغيرات معلومة.
4. **الارتباط المتعدد:** يستخدم الارتباط المتعدد لدراسة العلاقة بين المتغيرات الاحتمالية عندما يكون لدينا عدة متغيرات ونرغب بحساب الارتباط بين متغير محدد مع بقية المتغيرات.

تختلف العلاقة بين متغيرين من حيث قوّتها، فإذا كان تغير أحد هذين المتغيرين يعتمد كلياً على تغير الآخر نقول أن الارتباط بينهما كامل Perfect Correlation مثل العلاقة بين مساحة الدائرة ونصف قطرها، أمّا إذا كان الارتباط بين المتغيرات غير كامل أي أن تغير أحدهما لا يعتمد كلياً على تغير الآخر، فنقول بأن الارتباط هو ارتباط غير تام مثل العلاقة بين وزن الفرد وطوله، وبين التحصيل العلمي وعدد ساعات الدراسة.

معامل الارتباط r:

معامل الارتباط هو قيمة عددية تقيس شدة الارتباط ويُرمَز له بالرمز r إلّا إذا أراد الباحث أن يرمز لمعامل الارتباط برمز آخر غير r. ندرس معامل الارتباط لكي نحكم على الارتباط هل هو جيد أو سيء أو قوي أو ضعيف.

$$r \in [-1, +1]$$

فإذا كان: **r = +1** فإنّ العلاقة طردية لأن القيمة موجبة والارتباط موجب تام. أي أنّ الظاهرة الأولى عند زيادتها سوف تزيد الظاهرة الثانية وبنفس النسبة.

أمّا إذا كان: **r = -1** فإنّ العلاقة عكسية لأن القيمة سالبة والارتباط سالب تام. أي أنّ الظاهرة الأولى عند نقصانها سوف تزيد الظاهرة الثانية وبنفس النسبة. إنّ هاتين الحالتين السابقتين نادرتا الحدوث.

إذا كان: **r = 0** فإنّ العلاقة معدومة أي أن الارتباط معدوم ولا توجد علاقة بين المتغيرين المدروسين.

تتحدّد نوعية الارتباط بالاعتماد على الجدول (2) الآتي:

الجدول (2): أنواع الارتباط اعتماداً على قيم معامل الارتباط.

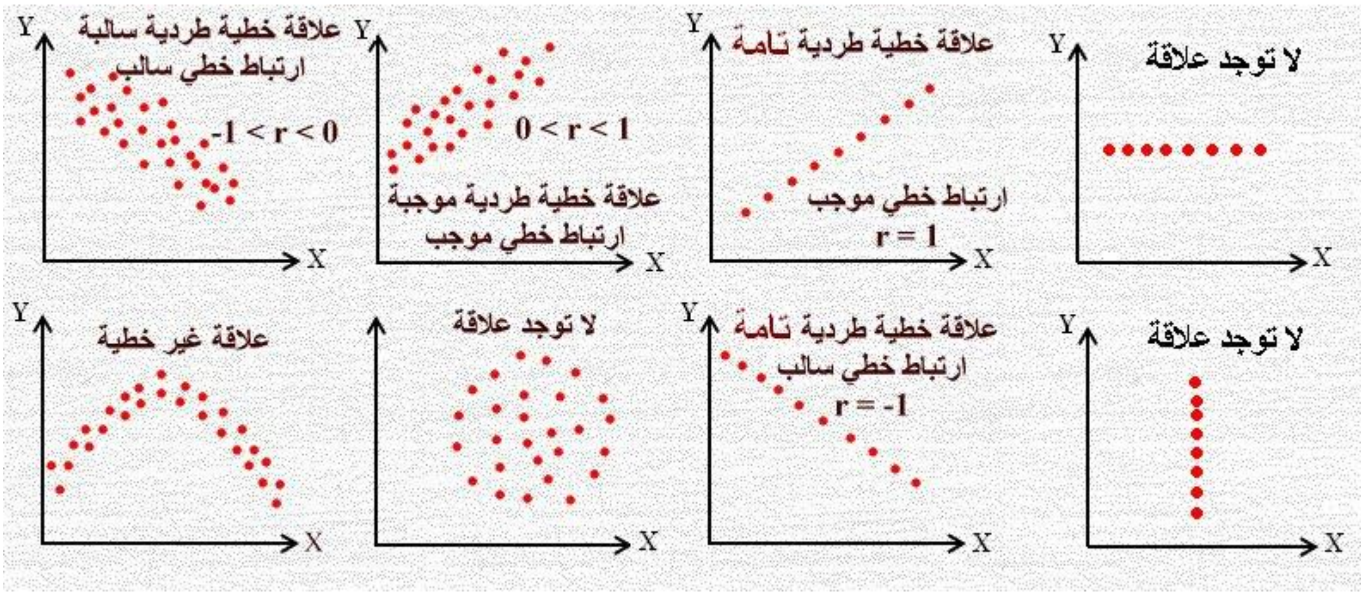
قيمة معامل الارتباط	نوع الارتباط
+1	ارتباط طردي تام
]1,7]	ارتباط طردي قوي
]7,4]	ارتباط طردي متوسط
]4,0[	ارتباط طردي ضعيف
0	ارتباط معدوم
]0,−0.4[	ارتباط عكسي ضعيف
[−0.4,−0.7[	ارتباط عكسي متوسط
[−0.7,−1[	ارتباط عكسي قوي
−1	ارتباط عكسي تام

الجدير بالذكر أن معامل الارتباط للظواهر الطبيعية بوجهٍ عام والزراعية بوجهٍ خاص لا يمكن أن تصل قيمته إلى الواحد بالقيمة المطلقة.

يجدر بالذكر أيضاً أن الارتباط البسيط قد يكون خطياً أو غير خطي، ويكون خطياً عندما يقترب شكل الانتشار الذي يمثّل قيم المتغيرين محلّ البحث من الخط المستقيم، وتكون قيمة معامل الارتباط أكبر كلما اقترب شكل الانتشار من الخط المستقيم.

يمكننا أن نحدّد شكل الارتباط بين ظاهرتين أو صفتين (y,x) اعتماداً على الرسم البياني في المستوي، كما هو مبيّن في المخطط (6) الآتي:

المخطط (6): أشكال الارتباط المختلفة بين ظاهرتين أو صفتين (y,x) اعتماداً على الرسم البياني في المستوي.



فإذا كان اتجاه الشكل الذي يحوي النقاط باتجاه الربع الأول وللأعلى بشكلٍ تام نقول أنّه لدينا ارتباط موجب حيث تزداد قيمة (y,x) وكلّما ضاق المجال كلّما كان الارتباط أقوى.

أمّا إذا كان اتجاه الشكل الذي يحوي النقاط باتجاه الربع الثاني وللأعلى بشكلٍ تام نقول أنّه لدينا ارتباط سالب حيث تتناقص قيمة (y,x) وكلّما ضاق المجال كلّما كان الارتباط أقوى.

حساب معامل الارتباط:

مهما تكن x وy فإنّ r يُحسَب من العلاقة الآتية:

$$r = \frac{\sum_{i=1}^{n}(xi-\bar{x}) * (yi-\bar{y})}{\sqrt{\sum_{i=1}^{n}(xi-\bar{x})^2 * \sum_{i=1}^{n}(yi-\bar{y})^2}}$$

تساؤل: ماذا نستفيد من معامل الارتباط؟

يفيدنا معامل الارتباط في حساب النسب المئوية التي تؤثّر بها إحدى الظاهرتين على الأخرى، وذلك عن طريق تربيع معامل الارتباط.

يُرمز لمربع معامل الارتباط بـ B. حيث أنّ: $B = r^2$ ويُسمَّى المعامل B **بمعامل التحديد.**

الانحدار Regression

معادلة الانحدار أو نموذج الانحدار يمكن تعريفه بأنّه علاقة رياضية تربط بين متحولين أو متغيرين أحدهما متغير تابع يكون في الطرف الأيسر من المعادلة والآخر يدعى متغير مستقل يكون في الطرف الأيمن من المعادلة.

يُعدُّ الانحدار أحد أهم الأساليب الإحصائية ويختصّ بقياس العلاقة بين متغير يسمّى المتغير التابع ومتغير آخر أو مجموعة متغيرات تسمّى المتغيرات المستقلة.

يجدر بالذكر أن نماذج الانحدار تُقسَم إلى:

أ. **انحدار خطي بسيط:** (والذي سنقوم بالتركيز عليه لأهميته وسهولة تقديره).

$$y = a + bx$$

ب. **انحدار خطي متعدد:**

$$y = a + b_1x_1 + b_2x_2 + b_3x_3$$

ج. **انحدار غير خطي بسيط:**

$$y = a + bx^2$$

د. **انحدار غير خطي متعدد:**

$$y = a + b_1x_1 + b_2{x_2}^2 + b_3{x_3}^3$$

عندما أريد دراسة ظاهرة أو صفة معينة (متغير تابع) كإنتاج نبات معيّن، بحيث يتأثر هذا الإنتاج بمجموعة من العوامل أو المتغيّرات المستقلّة المتفاوتة في تأثيرها على الإنتاج مثل كمية مياه الري، أو كمية السماد أو عدد العمال أو كمية الإصابات الحشرية والأمراض النباتية...إلخ. نقوم من أجل تسهيل الدراسة وتبسيطها بدراسة العلاقة بين الصفة المدروسة (المتغيّر التابع) وأهم العوامل المذكورة مثلاً (المتغيّر المستقل) وليكن كمية مياه الري على سبيل المثال. وبالتالي يمكن تمثيل ما سبق بالعلاقة الآتية:

$$y = a + bx$$

حيث تمثّل y كميّة الإنتاج من نبات معين، أمّا x فتمثّل كمية المياه المستخدمة في الري.

الانحدار الخطّي البسيط Simple Linear Regression:

هو أبسط أنواع الانحدار، والشكل العام لمعادلته (علاقة الانحدار الخطي): $y = a + bx$

y: الانحدار الخطي البسيط أو العامل أو المتغير. x : عامل مستقل.

a معامل ثابت (ثابت الانحدار): هو رقم ثابت، وهو الجزء المقطوع الذي يختلف حسب الدالّة.

مثلاً: في دالة الإنتاج: يمثّل الحدّ الأدنى من الإنتاج عندما تكون كمية المياه أو السماد المضاف تساوي الصفر.

في دالة العرض: يمثّل الحدّ الأدنى من الكمية المعروضة لو كان السعر صفراً.

في دالة الطلب: يمثّل الحد الأعلى من الكمية المطلوبة لو كان السعر صفراً.

b: معامل الانحدار (المعلمة الانحدارية): وتمثّل ميل الخط المستقيم الممثّل للعلاقة، وتمثّل المعامل الحدّي.

وتشير لمقدار التغيّر في المتغير التابع y نتيجة التغيّر بمقدار وحدة واحدة من المتغير المستقل x ، فكلما تغيّر العامل x بمقدار وحدة كاملة يتغيّر العامل y بمقدار وحدة كاملة، وقد يكون التغيّر إمّا بالزيادة أو بالنقصان.

عندما b **موجب** هذا يعني أن العامل المستقل يؤثر إيجاباً على العامل التابع.

عندما b **سالب** هذا يعني أن العامل المستقل يؤثر سلباً عل العامل التابع.

عند تتغيّر قيمة x بمقدار وحدة كاملة فإنّ قيمة y تتغير بمقدار b والذي هو معامل الانحدار.

علماً بأنّ:

$$a \in]-\infty, +\infty[$$

$$b \in]-\infty, +\infty[$$

أي أنّ a و b تأخذان جميع القيم السالبة والموجبة والكسرية والصحيحة.

قوانين معامل الانحدار b:

$$b = \frac{\sum_{i=1}^{n}(xi - \bar{x}) * (yi - \bar{y})}{\sum_{i=1}^{n}(xi - \bar{x})^2}$$

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

نستنتج من القانون الثاني أنّه لا يمكن حساب قيمة a إلّا بعد حساب قيمة b.

وبالتالي هناك ما ندعوه **تحليل الانحدار Regression Analysis**: هو كل طريقة إحصائية يتمّ فيها التنبّؤ بمتوسط متغير عشوائي أو عدّة متغيرات عشوائية اعتماداً على قيم وقياسات متغيرات عشوائية أخرى.

إنّ تحليل الانحدار هو أكثر من عملية ملائمة منحنى (أي اختيار المنحنى الأكثر ملائمة لمجموعة نقاط بيانية مُعطَاة) فهو يتضمّن ملائمة نموذج باستخدام مكوّنات حتمية واعتباطية (المكونات الحتمية تُدعَى المتنبئات أمّا المكونات الاعتباطية تُدعَى الخطأ).

الفرق بين الانحدار والارتباط:

من الارتباط نحصل على مؤشّر يصف العلاقة الخطية بين متغيرين ومدى قوّتها، أمّا في الانحدار يمكننا التنبّؤ بالعلاقة بين أكثر من متغيرين، ويمكن استخدام قيمة المتغير المستقل x لتوقع قيمة المتغير التابع y.

الارتباط يقيس مدى ارتباط متغيرين ولا يناسب خط، أمّا الانحدار الخطّي يجد الخط الأمثل الذي يتوقّع y من x.

مع الارتباط لا يهمّ التفكير في أي المتغيرين ندعوه x وأي المتغيرين ندعوه y، حيث ستحصل على معامل الارتباط نفسه إذا قمت بمبادلة الإثنين، أمّا في الانحدار فذلك يهمّ، حيث ستحصل على خط مختلف إذا بدّلت الإثنين.

كما أن الانحدار يفترض وجود علاقة سببية بين المتغيرين ويوضّح أيهما التابع وأيهما المستقل، أمّا الارتباط فيحدّد درجة اقتران المتغيرين دون أن يحدّد وجود أي علاقة تابعية أي أيهما تابع وأيهما مستقل.

ينحصر التشابه في الإشارة، حيث يتّفق معاملا الانحدار والارتباط بالإشارة فإذا كان أحدهما سالب فالآخر حتماً سالب، وإذ ساوى معامل الانحدار الصفر يكون معامل الارتباط مساوياً للصفر كذلك.

مثال: لدينا المعلومات التالية التي تبيّن عمر النبات مقدراً بالأسبوع وطول النبات مقدراً بالسنتيمتر.

والمطلوب: حساب كل من معامل الارتباط، ومعامل التحديد، وتشكيل معادلة الانحدار التي تعيّن العلاقة بين طول النبات وعمر النبات:

xi	yi	$xi-\bar{x}$	$yi-\bar{y}$	$(xi-\bar{x})*(\mathrm{yi}-\bar{\mathrm{y}})$	$(xi-\bar{x})^2$	$(yi-\bar{y})^2$
1	5	3-	19-	57	9	361
2	13	2-	11-	22	4	121
3	16	1-	8-	8	1	64
4	23	0	1-	0	0	1
5	33	1	9	9	1	81
6	38	2	14	28	4	196
7	40	3	16	48	9	256
28	168	0	0	172	28	1080
4	24					

حساب معامل الارتباط:

$$r=\frac{\sum_{i=1}^{n}(xi-\bar{x})*(yi-\bar{y})}{\sqrt{\sum_{i=1}^{n}(xi-\bar{x})^2*\sum_{i=1}^{n}(yi-\bar{y})^2}}$$

$$r=\frac{172}{\sqrt{28*1080}}=0.99$$

نستنتج أنّه لدينا علاقة طردية بين عمر النبات وطول النبات ولدينا ارتباط موجب قوي حيث أنّ زيادة عمر النبات تؤثّر وبشكل قوي على زيادة طول النبات.

في المثال السابق لدينا:

$$\mathrm{B}=r^2=0.99^2=0.98$$

نستنتج من ذلك أنّ:

98% من أسباب زيادة طول النبات تعود للزيادة في عمر النبات.

أمّا 2% من أسباب زيادة طول النبات تعود إلى أسباب أخرى غير مدروسة.

نقوم بتشكيل معادلة الانحدار التي تعيّن العلاقة بين الطول والعمر، كما يلي:

$$b = \frac{\sum_{i=1}^{n}(xi - \bar{x}) * (yi - \bar{y})}{\sum_{i=1}^{n}(xi - \bar{x})^2}$$

$$b = \frac{172}{28} = 6.14$$

بعد أن قمنا بحساب قيمة b نقوم بحساب قيمة a ، كما يلي:

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

$$a = 24 - 6.14 * 4 = -0.57$$

بعد استنتاج (a و b) نشكّل معادلة الانحدار:

$$y = -0.57 + 6.14x$$

b: موجب، وهذا يعني أنّ العامل المستقل x يؤثّر إيجاباً على العامل y أي أنّ زيادة العمر تؤدي إلى زيادة الطول.

كلما زاد العمر بمقدار وحدة كاملة يزداد طول النبات بمقدار وحدة كاملة والتي هي b=6.14 أو بكلام آخر نقول أنّ طول النبات يزداد أسبوعياً بمقدار 6.14 سم.

تساؤل: كم يزداد طول النبات إذا كان عمر النبات 4 أيام ؟

يزداد بمقدار يساوي $6.14 * \frac{4}{7} \approx 3.51$ سم

قد يُطلَب منّا أيضاً **اختبار معنوية معامل الارتباط:**

للقيام بهذا الاختبار نقارن T المحسوبة مع T الجدولية عند درجات حرية $n - 2$ ودرجة الخطأ أو مستويي المعنوية 1% و 5%.

فإذا كانت T المحسوبة > T الجدولية يعني هذا أن الارتباط بين x و y معنوي ومؤكّد إحصائياً أي غير ناتج عن الصدفة، ولا يمكن إهماله.

أمّا إذا كانت T المحسوبة < T الجدولية يعني أن الارتباط بين x و y غير معنوي وغير مؤكّد إحصائياً أي ناتج عن الصدفة، ويمكن إهماله.

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

T: T المحسوبة.

r: معامل الارتباط. r^2: معامل التحديد.

$n-2$: درجات الحرية حيث أن n تساوي عدد قيم x أو y.

إذا أردنا اختبار معنوية معامل الارتباط في مثالنا السابق علماً بأنّ قيمة T الجدولية عند درجة الخطأ 1% تساوي 4.32، نقوم بذلك كما يلي:

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

$$T = 0.99\sqrt{\frac{7-2}{1-0.98}} = 15.65$$

بعد مقارنة قيمة T المحسوبة مع T الجدولية نلاحظ أن T المحسوبة > T الجدولية عند درجة الخطأ أو مستوى المعنوية 1% ممّا يعني أنّ الارتباط بين x و y معنوي ومؤكّد إحصائياً أي غير ناتج عن الصدفة، وبالتالي لا يمكن إهماله.