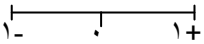
المحاضرة الثالثة عشر
تحليل الارتباط – ١

بالرغم من أن مقاييس العلاقة تختلف عما سبقها من مقاييس ، فهي تتعلق بدراسة العلاقة بين متغيرين (الإنتاجية والجودة) مثلا، بينما المقاييس السابقة فتهتم بدراسة الفروق بين المتغيرات .

وعندما نقول مقاييس العلاقة نعني بذلك تلك المقاييس التي تبين درجة العلاقة والارتباط بين متغيرين أو أكثر مثلا، كأن يكون الهدف معرفة هل هناك علاقة بين مستوى الإنتاجية وجودة المنتج في مصنع ما؟، أي هل كلما زادت الإنتاجية تقل جودة المنتج أو العكس .

معامل الارتباط: هو تعبير يشير إلى المقياس الإحصائي الذي يدل على مقدار العلاقة بين المتغيرات سلبية كانت أم إيجابية، وتتراوح قيمته بين الارتباط الموجب التام (+١) وبين الارتباط السالب التام (-١) .

((توضيح : أي ان معامل الإرتباط يستحيل ان يزيدعن+١ أويقل عن -١))

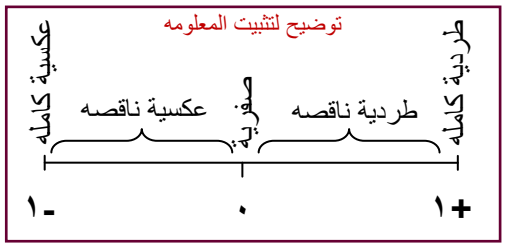


ومعظم الاحيان يأتي كسر ، واذا كان اعلى من+١ او اقل من-١ إذن راجع الحسابات ففيها خطأ

العلاقة الطردية بين المتغيرات: هو تعبير يشيرإلى تزايدالمتغيرين المستقل والتابع معا،فإذا كانت الإنتاجية مرتفعة، ومستوى الجودة مرتفع، يقال حينئذ أن بينهما ارتباط موجب، وأعلى درجة تمثله هي (+١).

العلاقةالعكسية بين المتغيرات: هوتعبيريشيرإلى تزايد في متغير يقابله تناقص في المتغير الآخر،فإذا كانت الإنتاجيةمنخفضة ومستوى الجودة مرتفع،يقال حينئذأن بينهماارتباط سالب،وأعلى درجةتمثلةهي (-١).

ومن الطبيعي ملاحظة أن الارتباط الكامل لا وجود له في الظواهر الطبيعية، وأن معامل الارتباط الناتج في الأبحاث والدراسات الإنسانية والاجتماعية يكون عادة كسرا موجبا أو سالبا. والجدول التالي يوضح أنواع العلاقات بين المتغيرات كما يصفها معامل الارتباط:



نوع العلاقة	قيمة معامل الارتباط
طردية كاملة	١+
طردية ناقصة	+ كسر(قيمة موجبة)
صفرية	صفر
عكسية ناقصة	- كسر (قيمة سالبة)
عكسية كاملة	١-

إن معامل الارتباط التام الموجب (+١) يعنى التغير في اتجاه واحد في كلا الظاهرتين مع بقاء الأوضاع النسبية لوحدات الظاهرة ثابتة، سواء كان هذا التغير في اتجاه الزيادة (أي زيادة قيم الظاهرة الأولى تتبعها زيادة في قيم الظاهرة الأخرى)، أو في اتجاه النقص (أي نقص قيم الظاهرة الأولى يتبعها نقص في قيم الظاهرة الأخرى) .

طرق التعرف على العلاقة بين متغيرين وحسابها:

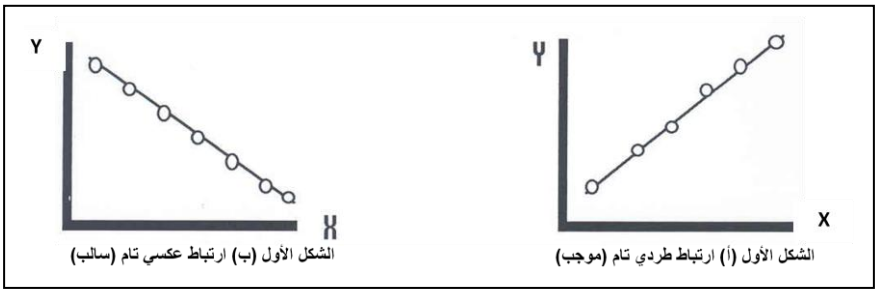
أولا: طريقة شكل الانتشار Scatter Diagram :

هناك وسيلة مبدئية يعرف الباحث من خلالها نوع الارتباط بين المتغيرين وما إذا كان الارتباط قوياً وضعيفاً أو منعدماً، وما إذا كانت العلاقة خطية أو غير خطية، موجبة أو سالبة. هذه الوسيلة هي " شكل الانتشار "والتي تصلح إذا كان المتغيران كميين. وجدير بالذكر أن هذه وسيلة مبدئية تساعد فقط في معرفة نوع الارتباط ولا تعتبربديلاًعن الطرق الإحصائيةالتي سوف نتناولهابالتفصيل في هذه المحاضرة.

والمقصود بشكل الانتشار هو تمثيل قيم الظاهرتين بيانياً على المحورين، المتغير الأول X على المحور الأفقي، والمتغير الثاني Y على المحور الرأسي، حيث يتم تمثيل كل زوج Pair من القيم بنقطة، فنحصل على شكل يمثل كيفية انتشار القيم على المستوى، وهو الذي يسمى شكل الانتشار. وطريقة انتشار القيم تدل على وجود أو عدم وجود علاقة بين المتغيرين ومدى قوتها ونوعها. فإذا كانت تتوزع بشكل منتظم دل ذلك على وجود علاقة (يمكن استنتاجها)، أما إذا كانت النقط مبعثرة ولا تنتشر حسب نظام معين دل ذلك على عدم وجود علاقة بين المتغيرين أو أن العلاقة بينهما ضعيفة. والأشكال التالية تظهر بعض أشكال الانتشار المعروفة :

الشكل الأول :

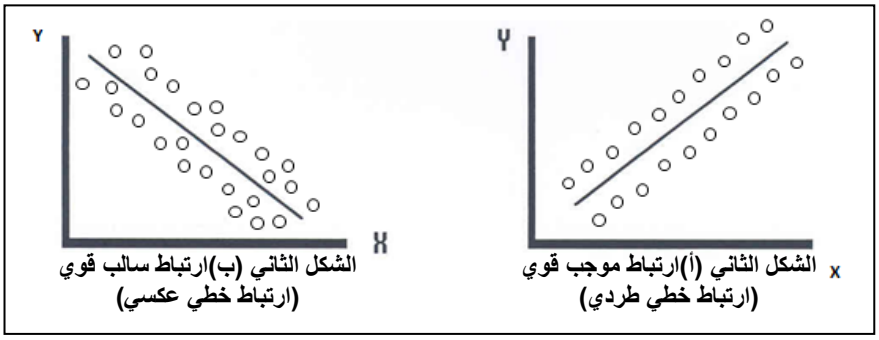
إذا وقعت جميع النقاط على خط مستقيم، دل ذلك على أن العلاقة بينهما خطية وأنها ثابتة أو تامة. وهذه تمثل أقوى أنواع الارتباط بين المتغيرين " ارتباط تام ". فإذا كانت العلاقة طردية فإن " الارتباط طردي تام " كما في الشكل الأول (أ). ومثاله العلاقة بين الكمية المشتراة من سلعة والمبلغ المدفوع لشراء هذه الكمية. أما إذا كانت العلاقة عكسية (وجميع النقاط تقع على خط مستقيم واحد فإن "الارتباط عكسي تام" كما في الشكل الأول (ب). ومثال على ذلك العلاقة بين السرعة والزمن.



الشكل الأول (ب) ارتباط عكسي تام (سالب)
الشكل الأول (أ) ارتباط طردي تام (موجب)

الشكل الثاني :

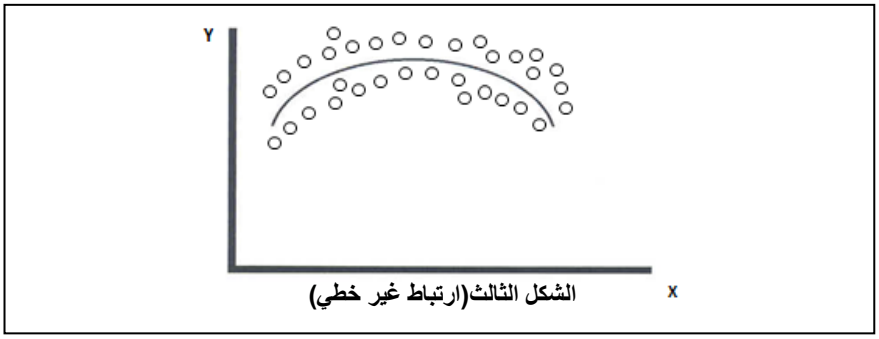
أما إذا كانت النقاط تأخذ شكل خط مستقيم ولكن لا تقع جميعها على الخط قيل أن العلاقة خطية (موجبة أو سالبة) كما في الشكل الثاني أ، ب. :



الشكل الثاني (ب)ارتباط سالب قوي (ارتباط خطي عكسي)
الشكل الثاني (أ)ارتباط موجب قوي (ارتباط خطي طردي)

الشكل الثالث :

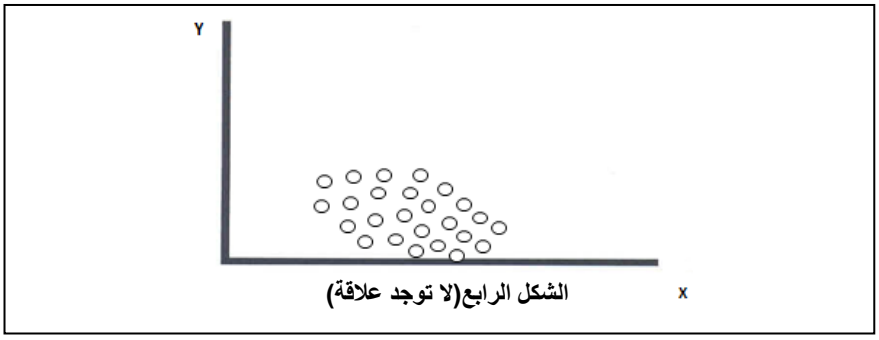
وإذا كانت العلاقة تأخذ شكل منحنى فإن الارتباط لا يكون خطياً "ارتباط غير خطي " Non Linear Correlation كما في الشكل الثالث :



الشكل الثالث(ارتباط غير خطي)

الشكل الرابع :

أما إذا كانت النقاط تتبعثر بدون نظام معين فإن ذلك يدل على عدم وجود علاقة بين المتغيرين (أو أن العلاقة بينهما ضعيفة جداً) كالعلاقة مثلاً بين دخل الشخص وطوله كما في الشكل الرابع :



الشكل الرابع(لا توجد علاقة)

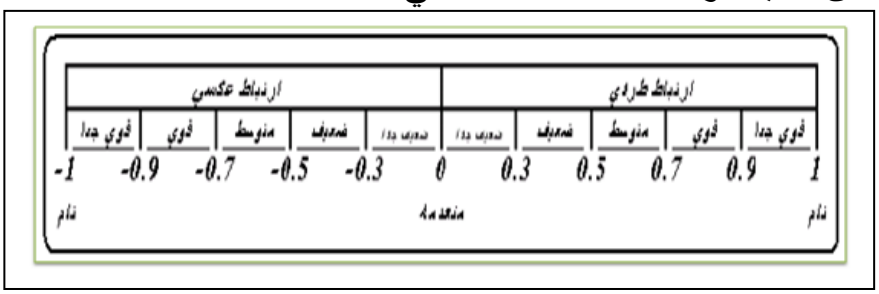
ثانيا: معامل الارتباط Correlation Coefficient :

يقاس الارتباط بين متغرين بمقياس إحصائي يسمى " معامل الارتباط " ويعكس هذا المقياس درجة أو قوة العلاقة بين المتغيرين واتجاه هذه العلاقة. وتنحصر قيمة معامل الارتباط بين + ١، - ١.

- فإذا كانت قيمة معامل الارتباط تساوي + ١ فمعنى ذلك أن الارتباط بين المتغيرين طردي تام، وهو أقوى أنواع الارتباط الطردي بين متغيرين.
- وإذا كانت قيمة معامل الارتباط تساوي – ١ فمعنى ذلك أن الارتباط بين المتغيرين عكسي تام، وهو أقوى أنواع الارتباط العكسي بين متغيرين.
- وإذا كانت قيمة معامل الارتباط تساوي صفر، فمعنى ذلك أنه لا يوجد ارتباط بين المتغيرين.
- وكلما اقتربت قيمة معامل الارتباط من + ١ أو – ١ كلما كان الارتباط قوياً، وكلما اقترب من الصفر كلما كان الارتباط ضعيفاً.

والخلاصة :

أنه كلما كانت العلاقة قوية بين المتغيرين كلما اقترب معامل الارتباط من + ١ أو – ١ فإذا وصلت قيمة المعامل إلى + ١ أو – ١ كان الارتباط تاماً بين المتغيرين. وأنه كلما كانت العلاقة ضعيفة بين المتغيرين كلما اقترب معامل الارتباط من الصفر، فإذا وصلت قيمة المعامل إلى الصفر كان الارتباط منعدماً بين المتغيرين. ومعنى ذلك أيضاً أنه لا يوجد ارتباط بين متغيرين تكون قيمة المعامل فيه أكبر من + ١ ولا أصغر من – ١. ويمكن تمثيل قوة العلاقة بالشكل التالي:



معامل بيرسون للارتباط الخطي البسيط Person's Correlation Coefficient

يعتبر معامل الارتباط الخطى البسيط لبيرسون Person's Correlation Coefficient والذى سنرمز له بالرمز r_p من أكثر الأدوات الإحصائية استخداما فى تحديد قوة العلاقة بين متغيرين كما يستعمل لتحديد مدى وجود علاقة خطية بين متغيرين. وهناك أكثر من صيغة يمكن الاعتماد عليها فى حساب معامل الارتباط الخطى البسيط لبيرسون منها:

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

وكذلك المعادلة الرياضية التالية والتي تعتبر اسهل وابسط: (من خلال الدرجات الخام)

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

وتتراوح قيمة معامل الارتباط بين الواحد الصحيح الموجب و الواحد الصحيح السالب أى أن قيمة معامل تكون كالتالي: $1 \geq r_p \geq -1$ **والارتباط غالبا قيمته كسر أي اقل من الواحد الصحيح.**

ولتحديد نوع العلاقة نعتمد على اشارة معامل الارتباط فإذا كانت الإشارة:

- موجبة فإن العلاقة تكون طرديه
- سالبة فإن العلاقة تكون عكسية

ولتحديد قوة العلاقة نعتمد على قيمة معامل الارتباط فإذا كانت القيمة:

- من أكبر من صفر إلى أقل من 0.3 فتكون علاقة ضعيفة جدا
- من أكبر من 0.3 إلى أقل من 0.5 تكون علاقة ضعيفة
- من أكبر من 0.5 إلى أقل من 0.7 تكون علاقة متوسطة
- من أكبر من 0.7 إلى أقل من 0.9 تكون علاقة قوية
- من أكبر من 0.9 إلى أقل من 1.00 تكون علاقة قوية جدا
- الواحد الصحيح تكون علاقة تامة
- إما إذا كانت قيمة معامل الارتباط تساوى صفر فلا توجد علاقة خطيه او ارتباط بينهما أى يكون المتغيرين مستقلين عن بعضهما البعض وتكون العلاقة منعدمة .

فمثلا إذا كانت قيمة معامل الارتباط r_p كالتالي فإن تفسيرة يكون:

قيمة	تفسير معامل الارتباط
0.91	أرتباط طردى قوى جدا
-0.87	ارتباط عكسى قوى
-0.21	ارتباط عكسى ضعيف جدا
0.43	ارتباط طردى ضعيف
1	ارتباط طردى تام
-0.51	ارتباط عكسى متوسط

مثال: فيما يلى بيان بالمنفق على الاعلان والمبيعات لأحد المنتجات فكانت بالمليون ريال كمايلى:

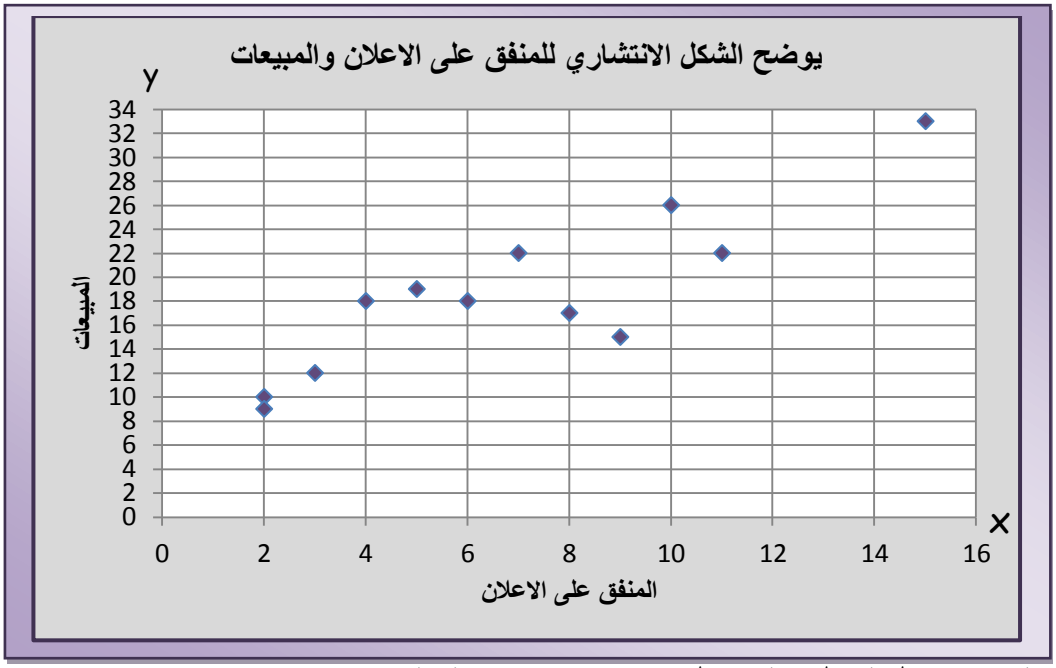
المنفق على الاعلان	2	3	12	7	22	6	18	5	19	10	26	15	33	4	18	11	22	9	15	8	17
المبيعات	10	12	9	22	18	18	19	26	10	5	33	18	22	4	11	9	15	17			

المطلوب:

ارسم شكل الانتشار يوضح العلاقة بين المنفق على الاعلان و المبيعات ؟

احسب معامل الارتباط الخطى البسيط (بيرسون)، مع التعليق

الحل: ١-ارسم شكل الإنتشار يوضح العلاقة بين المنفق على الاعلان والمبيعات؟



٢-احسب معامل الارتباط الخطى البسيط (بيرسون)، مع التعليق

بتطبيق المعادلة: $r_p = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$ وذلك يحتاج عمل جدول لتطبيق المعادلة وحساب الإرتباط:

x	y	xy	x^2	y^2
٢	١٠	٢٠	٤	١٠٠
٣	١٢	٣٦	٩	١٤٤
٢	٩	١٨	٤	٨١
٧	٢٢	١٥٤	٤٩	٤٨٤
٦	١٨	١٠٨	٣٦	٣٢٤
٥	١٩	٩٥	٣٥	٣٦١
١٠	٢٦	٢٦٠	١٠٠	٦٧٦
١٥	٣٣	٤٩٥	٢٢٥	١٠٨٩
٤	١٨	٧٢	١٦	٣٢٤
١١	٢٢	٢٤٢	١٢١	٤٨٤
٩	١٥	١٣٥	٨١	٢٢٥
٨	١٧	١٣٦	٦٤	٢٨٩
٨٢	٢٢١	١٧٧١	٧٣٤	٤٥٨١

ثم يمكن تطبيق المعادلة بعد الجدول كمايلي:

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

$$= \frac{12\times(1771) - (82\times 221)}{\sqrt{12\times(734) - (82)^2}\times\sqrt{12\times(4581) - (221)^2}}$$

$$= \frac{21252 - 18122}{\sqrt{8808 - 6724}\times\sqrt{54972 - 48841}}$$

$$= \frac{3130}{\sqrt{2084}\times\sqrt{6131}}$$

$$= \frac{3130}{45.6508\times 78.3007}$$

$$= \frac{3130}{3574.4895} = +0.87564$$

دل ذلك وجودعلاقة طردية قوية بين المنفق على الاعلان والمبيعات

سمر المغربي

ومن أهم خصائص معامل الارتباط الخطى البسيط لبيرسون أنه لا يعتمد على قيم المتغيران نفسها عند حساب قيمته وإنما يعتمد على مقدار التباعد بين هذه القيم بعضها البعض.
لذلك لا يتأثر معامل الارتباط الخطى البسيط بأى عمليات جبرية يتم إجراءها على بيانات اى من المتغيرين أو أحدهما من جمع أو طرح أو ضرب أو قسمة.

مثال: فى بيانات المثال إذا أكتشفت إدارة الشركة أن البيانات تم تجميعها وحسابها بطريقة خاطئة حيث يجب إضافة ٥ مليون ريال إلى جميع قيم المنفق على الإعلان. كما أن المبيعات يجب مضاعفة قيمتها لجميع القيم.

المنفق على الاعلان	2	3	12	7	22	9	18	6	19	5	26	10	33	15	18	4	22	11	15	9	8
المبيعات	10	12	9	22	18	33	26	19	18	22	15	17									

المطلوب:
أحسب معامل الارتباط فى هذه الحالة بين المنفق على الإعلان والمبيعات.

الحل:

بتطبيق المعادلة: $r_p = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$ وذلك يحتاج عمل جدول لتطبيق المعادلة وحساب الإرتباط:

x	y	xy	x^2	y^2
٧	٢٠	١٤٠	٤٩	٤٠٠
٨	٢٤	١٩٢	٦٤	٥٧٦
٧	١٨	١٢٦	٤٩	٣٢٤
١٢	٤٤	٥٢٨	١٤٤	١٩٣٦
١١	٣٦	٣٩٦	١٢١	١٢٩٦
١٠	٣٨	٣٨٠	١٠٠	١٤٤٤
١٥	٥٢	٧٨٠	٢٢٥	٢٧٠٤
٢٠	٦٦	١٣٢٠	٤٠٠	٤٣٥٦
٩	٣٦	٣٢٤	٨١	١٢٩٦
١٦	٤٤	٧٠٤	٢٥٦	١٩٣٦
١٤	٣٠	٤٢٠	١٩٦	٩٠٠
١٣	٣٤	٤٤٢	١٦٩	١١٥٦
١٤٢	٤٤٢	٥٧٥٢	١٨٥٤	١٨٣٢٤

ثم يمكن تطبيق المعادلة بعد الجدول كمايلي:

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

$$= \frac{12\times(5752) - (142\times 442)}{\sqrt{12\times(1854) - (142)^2}\times\sqrt{12\times(18324) - (442)^2}}$$

$$= \frac{69024 - 62764}{\sqrt{22248 - 20164}\times\sqrt{219888 - 195364}}$$

$$= \frac{6260}{\sqrt{2084}\times\sqrt{24524}}$$

$$= \frac{6260}{45.6508\times 156.6014}$$

$$= \frac{6260}{7149.2530} = +0.87564$$

وهي نفس النتيجة التي حصلنا عليها سابقاً مما يدل على أن معامل الإرتباط لم تتأثر قيمته بالعمليات الجبرية من جمع (٥ مليون) أوالضرب(×٢) وبالمثل لايتأثر بالطرح أو القسمة.

معامل التحديد Determination Coefficient

وهو مربع معامل الارتباط لذلك يرمز له بالرمز R^2 أو R-Square و هو يشير إلى نسبة تفسير المتغير أو المتغيرات المستقلة للتغير فى المتغير التابع

فمثلا: نجد أن المنفق على الاعلان يفسر نسبة (0.8756^2) أى 76.675 % من التغير فى قيمة المبيعات بينما 23.32 % من التغير فى المبيعات ترجع إلى عوامل أخرى منها الخطاء العشوائى .