الانحدار والارتباط

أولا - تحليل الانحدار

يعتبر الانحدار من المواضيع الأساسية في النظرية الإحصائية بشكل عام، حيث يستخدم على نطاق واسع في العديد من العلوم الطبيعية او الإدارية أو الاقتصادية، وذلك لاستكشاف القوانين التي تربط الظواهر فيما بينها، ثم وضع هذه القوانين في شكل نماذج رياضية يطلق عليها "**نماذج الانحدار**" تُسهّل عملية تقدير قيم هذه الظواهر بدلالة قيم الظواهر الأخرى.

ويعد العالم الإحصائي الانجليزي Francis Galton أول من استخدم مفهوم الانحدار في التطبيقات البيولوجية تحديدا، وذلك لاستكشاف بعض العلاقات التي تربط بعض المتغيرات البيولوجية.

يطلق على الظاهرة المراد تقدير قيمتها -وبالتالي معرفه سلوكها- **بالمتغير التابع** أو **المتغير المفسَّر**، ويرمز لها عادة بالرمز Y ، وبالمقابل يطلق على المتغير الآخر أو بقية المتغيرات الأخرى المتحكمة فيها **بالمتغير المستقل** او **المتغير المفسِّر**، ويرمز له عادة بالرمز X.

وبشكل عام تعطينا دراسة الانحدار إجابة عن سؤالين رئيسين:

- ✓ **الأول**: هل توجد علاقة بين المتغيرين المدروسين؟
- ✓ **الثاني**: إذا وجدت علاقة بين هذين المتغيرين (أو أكثر)، فما هو الشكل الرياضي لهذه العلاقة؟

يمكن تقسيم الانحدار من حيث شكلُ العلاقة الى قسمين مهمين: انحدار خطي وانحدار غير خطي.

أما الانحدار الخطي فتكون فيه العلاقة "خطية" بين الظاهرة المدروسة من جهة والمتغيرات المتحكمة فيها من جهة أخرى، ومعنى ذلك أن أي تغير يحصلُ في المتغير المستقل يتبعه مباشرة تغير ثابت في المتغير التابع.

وأما الانحدار غير الخطي فتكون فيه العلاقة "غير خطية" بين الظاهرة المدروسة والمتغيرات المؤثرة فيها كأن تكون علاقة أسية أو علاقة قطع مكافئ أو قطعا زائدا... الخ. وسنركز في محاضرتنا على القسم الأول من الانحدار.

الانحدار الخطي البسيط :إن الهدف الرئيس من تحليل الانحدار الخطي عموما هو تقدير العلاقة الرياضية الخطية التي تربط بين متغيرين اثنين أو أكثر، وبهذا الصدد يمكن أن نجد نوعين من الانحدار الخطي: بسيط ومتعدد.

أما **الانحدار الخطي البسيط** فهو الذي يدرس العلاقة بين "متغيرين اثنين" فقط، أحدهما تابع والآخر مستقل، وأما **الانحدار الخطي المتعدد** فهو الذي يدرس العلاقة الخطية بين "عدد من المتغيرات"، أحدها تابع والبقية متغيرات مستقلة. وفي هذه المحاضرة سنصب اهتمامنا على "**الانحدار الخطي البسيط**"، أين نقوم من خلاله بتقدير العلاقة الخطية بين متغيرين فقط -كما ذكرنا- أحدهما مستقل والآخر تابع. يمكن تلخيص النموذج الرياضي الخطي للانحدار الخطي البسيط في المعادلة الأتية :

$$\hat{Y} = aX + b$$

وتسمى معادلة انحدار Y على X ، حيث:

$\hat{Y}$ القيم التقديرية للمتغير التابع Y.

X المتغير المستقل.

a هو ميل خط الانحدار، ويمثل مقدار التغير المقدَّر في Y إذا تغير X بوحدة واحدة.

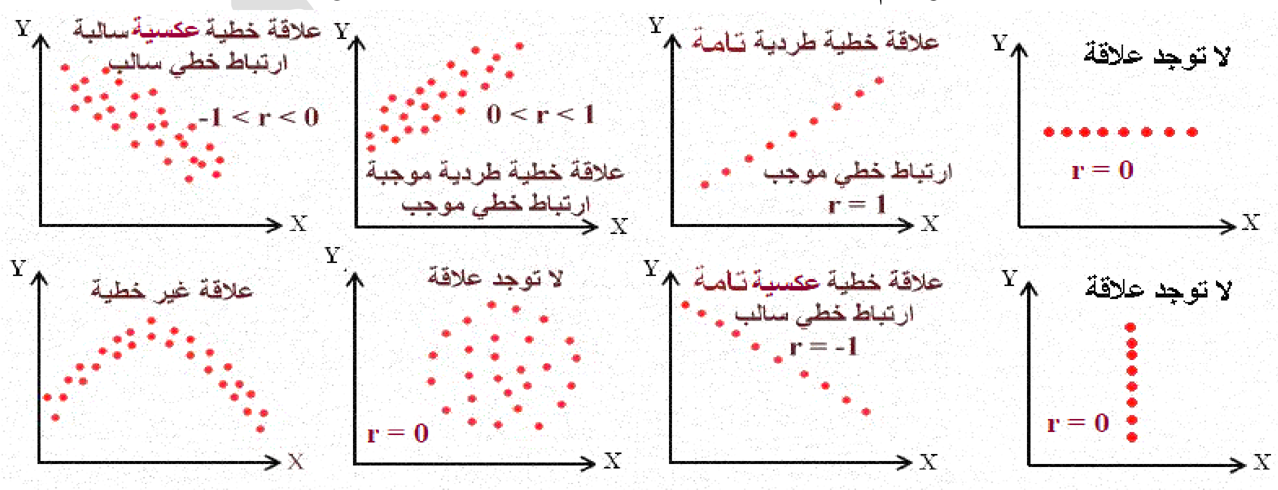
b هو القيمة المقدرة للمتغير Y إذا انعدم المتغير X.

وهذان الأخيران هما قيمتان ثابتتان تمثلان "**معلمتي**" النموذج، وهما المعلمتان الواجب تحديدهما ليتكشَّف لنا النموذج كاملا.

✓ **طرق تقدير معالم نموذج الانحدار الخطي البسيط:** إن أول خطوة يعمدُ إليها الباحث لتحديد معلمتي نموذج الانحدار هي رسمه لما يسمى **بلوحة الانتشار** أو **سحابة النقاط.** والتي تنتج عن تعيين مختلف الثنائيات (x , y) في معلم متعامد، يوضع في محوره الأفقي المتغير المستقل X وفي محور العمودي المتغير التابع Y ، وبحسب شكل هذه السحابة يمكن أن نحدد طبيعة العلاقة بين المتغيرين المستقل والتابع:

فإذا ظهر شكل لوحة الانتشار يقتربُ من تشكيل **خط مستقيم** نقول إن العلاقة خطية بين المتغيرين، وإذا اخذت السحابة شكلا آخر غير الشكل المستقيم، كأن تكون أسيةً مثلا، فنقول إن العلاقة أسية، وإذا اخذت **شكلا عشوائيا** لا يشبه أي شكل رياضي معروف، نقول إنه لا توجد علاقة أصلا بين المتغيرين المدروسين. (أنظر الشكل 01 أسفله)

الشكل رقم 01: سحابة النقاط وأشكال العلاقة بين المتغيرين X و Y.



المصدر: الرابط: https://www.almuheet.net/post/247336

بالنسبة للانحدار الخطي البسيط فإنه يُنتظر -بعد رسم لوحة الانتشار- أن نحصُل على سحابة تقترب من تشكيل خط مستقيم، ويتوقف مدى انتشار هذه السحابة وتشتتها على قوة العلاقة بين المتغيرين؛ حيث كلما قويت هذه العلاقة قلَّ تشتت هذه السحابة، واقتربت من الوقوع على خط مستقيم، وعندما يحدث ذلك (تقع جميع النقاط على خط مستقيم) تعتبر حينها العلاقة تامة أو وظيفية، والعكس صحيح، والأشكال السابقة تبين ذلك بوضوح.

لكن السؤال المطروح هنا: **كيف نحدد بدقة الخط المستقيم الذي تقتربُ السحابةُ من تشكيله؟** وهذا لتحديد معلمتي نموذج الانحدار a و b السابقتين؟ للإجابة عن هذا السؤال هناك طريقتان:

- **الطريقة الأولى - الطريقة البيانية**: أو التحديد بالنظر؛ حيث يقوم الباحث بوضع أقرب خط يراه مناسبا لتمثيل السحابة، وهو الخط الأقرب الى جميع النقاط كما يبدو له. لكن يعاب على هذه الطريقة أنها غير دقيقة، حيث انه يمكن لعدد من الاشخاص ان يرسموا خطوطا مختلفة للسحابة نفسها، بل إن الشخص نفسه يمكن ان يرسم عدة خطوط في مناسبات مختلفة.
- **الطريقة الثانية - طريقه المربعات الصغرى**: وهي أشهر الطرق وأدقها لتحديد معالم نموذج الانحدار، تقوم فكرتها الأساسية على جعل "المسافات" او "الانحرافات" بين سحابة النقاط (القيم الحقيقية للمتغير التابع Y) وبين خط الانحدار (القيم التقديرية للمتغير التابع $\hat{Y}$) "**أقل ما يمكن**."

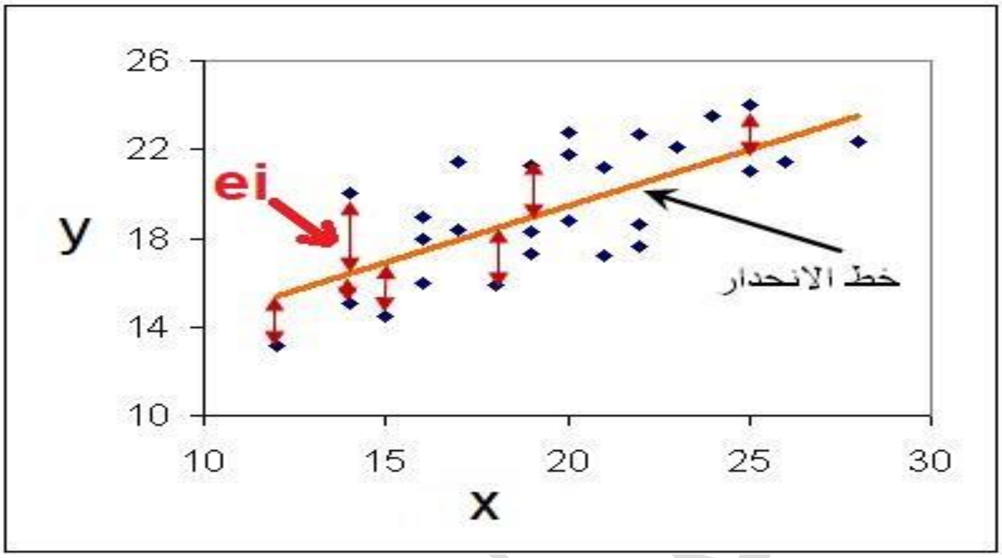
ولأن أفضل خط نبحث عنه يؤدي دورا في سحابة النقاط يشبه دور الوسط الحسابي في مجموعة القيم، فإن مجموع انحرافات القيم الحقيقية Y في السحابة عن القيم المقدرة $\hat{Y}$ الواقعة على الخط يساوي "**الصفر**" حسب الخاصية رقم 01 من خصائص الوسط الحسابي التي رأيناها سابقا.

نرمز لهذه الانحرافات بالرمز e_i حيث عند $\sum e_i = \sum(Y_i - \hat{Y}_i) = 0$

ومعنى هذا أن مجموع "**مربعات**" هذه الانحرافات يجب أن يكون أصغر ما يمكن عند أفضل خط نبحث عنه، وهذا حسب الخاصية رقم 02 من خصائص الوسط الحسابي التي رأيناها سابقا، وهذا هو سبب تسميتها بطريقة المربعات الصغرى، حيث أن مجموع e_i^2 هو دالة في المعلمتين a و b. (أنظر الشكل 02 أسفله)

$$\sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - ax - b)^2 = f(a, b)$$

الشكل رقم **02**: توضح انحراف القيمة الحقيقية عن القيمة المقدَّرة.



المصدر: https://samehar.wordpress.com/2009/11/11/120909/

لتكون الدالة f أصغر ما يمكن يجب أن يتوافر فيها شرطان: الأول لازم والثاني كافٍ.

- ✓ **الشرط اللازم**: يجب أن تنعدم المشتقتان الجزئيتان الأوليان بالنسبة للمعلمتين a و b.
- ✓ **الشرط الكافي**: أن تكون المشتقتان الجزئيتان الثانيتان بالنسبة للمعلمتين a و b موجبتين.
- ✓ **الشرط اللازم**:
- ➢ نشتق f بالنسبة للمعلمة a ونضعها تساوي الصفر:

$$\left(\frac{\delta \sum e_i^2}{\delta a} = 0\right) \Leftrightarrow \left[-2\sum (Y_i - aX_i - b)X_i = 0\right] \Leftrightarrow \left[\sum (Y_i - aX_i - b)X_i = 0\right]$$

$$\Leftrightarrow \left[\sum (X_iY_i - aX_i{}^2 - bX_i) = 0\right] \Leftrightarrow \left(\sum X_iY_i - a\sum X_i{}^2 - b\sum X_i\right) = 0$$

$$\Leftrightarrow \sum X_iY_i = a\sum X_i{}^2 + b\sum X_i \ldots\ldots\ldots\ldots\ldots.(1)$$

- ➢ نشتق f بالنسبة للمعلمة b ونضعها تساوي الصفر:

$$\left(\frac{\delta \sum e_i^2}{\delta b} = 0\right) \Leftrightarrow \left[-2\sum (Y_i - aX_i - b) = 0\right] \Leftrightarrow \left[\sum (Y_i - aX_i - b) = 0\right]$$

$$\Leftrightarrow \left(\sum Y_i - a\sum X_i - nb\right) = 0 \Leftrightarrow \sum Y_i = a\sum X_i + nb \ldots\ldots\ldots(2)$$

من (1) و (2) يمكن أن نُكوِّن جملة معادلتين ذات مجهولين a و b كما يلي:

$$\begin{cases} \sum X_iY_i = a\sum X_i^{\ 2} + b\sum X_i \\ \sum Y_i = a\sum X_i + nb \end{cases}$$

وبحل هذه الجملة -بطريقة الجمع أو التعويض المعروفتين- نخلص إلى أن:

$$a = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{n\sum X_i^{\ 2} - (\sum X_i)^2}$$

$$b = \frac{\sum X_i^{\ 2}\sum Y_i - \sum X_i \sum X_iY_i}{n\sum X_i^{\ 2} - (\sum X_i)^2} = \frac{1}{n}\left[\sum Y_i - a\sum X_i\right] = \bar{Y} - a\bar{X}$$

مثال: يبين الجدول التالي البينات الخاصة بعينة عشوائية من منتجي الأدوية، حيث X تمثل مصاريف البحث، و Y تمثل الأرباح (بمليون دينار):

Y	X
2	1
5	3
9	5
10	7
14	9
40	**25**

1. أرسم لوحة انتشار البيانات.
2. ما هو شكل العلاقة بين X و Y؟
3. أوجد معادلة خط الانحدار.
4. ما ذا تعني قيمة الثابتين a و b؟
5. حدد القيمة التقديرية للربح إذا كانت المصاريف 10 مليون دينار.

الحل:

1. رسم لوحة الانتشار (سحابة النقاط) يكون في معلم متعامد أين نحدد عليه الثنائيات (x_i, y_i).
2. بملاحظة لوحة الانتشار نلاحظ أن النقاط تقترب من تشكيل خط مستقيم، وعليه فشكل العلاقة بين المتغيرين خطية شكلها الرياضي: $\hat{Y} = aX + b$
3. تحديد المعادلة يعني تحديد القيمتين a و b

X^2	XY	Y	X
1	2	2	1
9	15	5	3
25	45	9	5
49	70	10	7
81	126	14	9
165	**258**	**40**	**25**

$$a = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{n\sum X_i^{\ 2} - (\sum X_i)^2} = \frac{5(258) - 25 \times 40}{5(165) - 25^2} = \mathbf{1.45}$$

$$b = \frac{\sum X_i^{\ 2}\sum Y_i - \sum X_i \sum X_iY_i}{n\sum X_i^{\ 2} - (\sum X_i)^2} = \frac{165 \times 40 - 25 \times 258}{5(165) - 25^2} = \mathbf{0.75}$$

ومنه المعادلة كالتالي: $\hat{\boldsymbol{Y}} = \mathbf{1.45}\boldsymbol{X} + \mathbf{0.75}$

4. تمثل قيمة a "**الميل**"، ومعناها مقدار الزيادة "**التقديرية**" في الأرباح كلما صرفنا وحدة واحدة من المصاريف، أي بعبارة أخرى أن الأرباح يُقدر ان تزيد بمقدار 1450000 دج كلما زدنا مقدار المصاريف بمبلغ 1000000 دج.

أما b فقيمتها تمثل مقدار الأرباح "**المقدَّرة**" إذا لم نصرف أيَّ دينار في البحث والتطوير، يقدر هذا الربح بمبلغ 750000 دج. (لأن وحدة القياس في التمرين هي مليون دينار)

5. تحديد القيمة التقديرية للربح إذا كانت المصاريف 10 مليون دينار: يتم ذلك بالتعويض مباشرة في المعادلة.

$$\hat{Y} = 1.45(10) + 0.75 = 15.25$$

أي أننا لو صرفنا 10 مليون دينار كمصاريف بحث وتطوير فسنجني أرباحا متوقعة تقدر بمبلغ 15250000 دج.

- ✓ **الخطأ المعياري للتقدير $S_{Y/X}$**:

يعبر هذا الخطأ عن الانحراف (التشتت) بين القيم التقديرية $\hat{Y}$ التي مصدرها المعادلة، وبين القيم الحقيقية Y التي مصدرها الواقع، وهو يشبه كثيرا -في حسابه- فكرة الانحراف المعياري الذي درسناه سابقا، يمكن حساب الخطا المعياري للتقدير وفق القانن الآتي:

$$S_{Y/X} = \sqrt{\frac{\sum\left(Y_i - \hat{Y}_i\right)^2}{n}}$$

- ✓ **معادلة انحدار X على Y**:

درسنا في الجزء السابق من الدرس انحدار Y على X ، حيث اعتبرنا المتغير X مستقلا عن Y ومؤثرا فيه. ولأن العلاقات بين الظواهر كثيرا ما تكون متبادلة التأثير (كل من المتغيرين يؤثر في الآخر ويتأثر به)، فإنه من الممكن دراسة انحدار X على Y أين يكون المتغير X تابعا والمتغير Y مستقلا. ويكون شكلها على النحو الآتي:

$$\hat{X} = \grave{a}Y + \grave{b}$$

يتم ذلك بالطريقة نفسها التي تطرقنا إليها لدى دراستنا لانحدار Y على X أين نجد:

$$\grave{a} = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{n\sum {Y_i}^2 - (\sum Y_i)^2}$$

$$\grave{b} = \frac{\sum {Y_i}^2 \sum X_i - \sum Y_i \sum X_iY_i}{n\sum {Y_i}^2 - (\sum Y_i)^2} = \frac{1}{n}\left[\sum X_i - \grave{a}\sum Y_i\right] = \bar{X} - \grave{a}\bar{Y}$$

ملاحظة: إذا كان انحدار Y على X خطيا فإن انحدار X على Y يكون خطيا أيضا، وإذا كان انحدار Y على X طرديا فإن انحدار X على Y يكون طرديا أيضا، والعكس صحيح.

الانحدار والارتباط

ثانيا - تحليل الارتباط.

يهتم تحليل الارتباط بقوة العلاقة بين متغيرين أو أكثر، مثل قوة العلاقة بين مهارة العاملين والأنتاجية، أو بين سعر السلعة والكمية المطلوبة منها... حيث يجيبنا تحليل الارتباط عن سؤال رئيس هو: **ما شدة (قوة) العلاقة بين المتغيرين X و Y؟**

تقاس قوة العلاقة هذه بمؤشر إحصائي يدعى معامل الارتباط، يرمز له عادة بالرمز r، حيث يمكن أن نجد عدة أنواع من معاملات الارتباط، تختلف باختلاف عدد المتغيرات المدروسة وطبيعتها، أهم هذه المعاملات:

- ✓ معامل بيرسون للارتباط الخطي البسيط r_p.
- ✓ معامل ارتباط الرتب (معامل سبيرمان) r_s.
- ✓ معامل الاقتران r_a.
- ✓ معامل التوافق r_c.

أ. معامل بيرسون للارتباط الخطي البسيط r_p:

1. **تعريفه**: هو عبارة عن مؤشر يحدد قوة العلاقة الخطية واتجاهها بين متغيرين اثنين، وهو مستقل عن وحدة القياس التي تقاس بها الظاهرة Y أو المتغير X. يمكن إيجاده عن طريق حساب الوسط الهندسي للميلين a و a' لمعادلتي الانحدار السابقتين. أي أن:

$$r_p = \sqrt{a \times a'}$$

يمكن حساب هذا المعامل بقوانين أخرى أهمها:

$$r_p = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{\sqrt{\left[n\sum {X_i}^2 - (\sum X_i)^2\right] \times \left[n\sum {Y_i}^2 - (\sum Y_i)^2\right]}}$$

وهو الأسهل تطبيقا على الآلة الحاسبة

$$= \frac{\sum X_iY_i - n\bar{X}\bar{Y}}{\sqrt{\left[\sum {X_i}^2 - n(\bar{X})^2\right]\left[\sum {Y_i}^2 - n(\bar{Y})^2\right]}}$$

2. **خصائصه**: يختص معامل بيرسون للارتباط الخطي بما يلي:
   - ✓ قيمته محضورة في المجال المغلق من (-1) إلى (+1).
   - ✓ كلما اقتربت قيمته من طرفي المجال ازداد الارتباط قوة، وكلما اقتربت قيمته من الصفر ضعف الارتباط.
   - ✓ تدل قيمته الموجبة على أن الارتباط طردي، وتدل قيمته السالبة على أن الارتباط عكسي.
   - ✓ تدل قيمته المعدومة على عدم وجود ارتباط بين المتغيرين.

- ✓ يكون الارتباط تماما (والعلاقة وظيفية) إذا كانت قيمة معامل الارتباط تساوي (-1) أو (+1)، وهذا لا يكون في الظواهر الإحصائية (الظواهر الاقتصادية والاجتماعية والتربوية ...)

3. **مثال توضيحي**: تمثل البيانات التالية عدد أيام الغياب X لخمسة طلبة مقابل نقاطهم في أحد اختبارات الرياضيات Y . **المطلوب**: بافتراض أن العلاقة بين المتغيرين خطية: أحسب معامل بيرسون للارتباط.

Y^2	X^2	XY	النقطة Y	أيام الغياب X
729	0	0	27	0
441	16	84	21	4
324	16	72	18	4
100	100	100	10	10
256	36	96	16	6
1850	**168**	**352**	**92**	**24**

$$r_p = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{\sqrt{\left[n\sum {X_i}^2 - (\sum X_i)^2\right] \times \left[n\sum {Y_i}^2 - (\sum Y_i)^2\right]}}$$

$$= \frac{5(352) - (24)(92)}{\sqrt{[5(168) - (24)^2] \times [5(1850) - (92)^2]}}$$

$$= (-0.98) = -98\%$$

ومعنى هذا أن هناك ارتباطا عكسيا قويا جدا بين عدد أيام الغياب والنقطة في امتحان الرياضيات، حيث كلما زاد عدد أيان الغياب نقصت النقطة.

4. **تحديد درجة الثقة**: يمكن تحديد درجة الثقة في تقديرات معادلة الانحدار من خلال حساب ما يسمى بمعامل التحديد، والذي يرمز له بالرمز R^2، وهو معامل تتراوح قيمته بين 0 و 1 ، حيث يمكن ان يعبر عنه في شكل نسبة مئوية، وهو عبارة عن مربع معامل ارتباط بيرسون. أي أن:

$$R^2 = \left(r_p\right)^2$$

يمكن أيضا حساب معامل التحديد باستخدام ما يسمى "**مجموع الانحرافات المربعة**"، التي يمكن تصنيفها إلى ثلاثة أنواع من الانحرافات:

- ✓ **الانحرافات المفسَّرة**: وهي الانحرافات بين القيم التقديرية $\hat{Y}$ (التي مصدرها معادلة الانحدار) وبين قيمة الوسط الحسابي $\bar{Y}$ (الوسط الحسابي للقيم الحقيقية Y_i). وسميت كذلك لأن هذا النوع من الانحرافات مفسَّر بالانحدار.

يرمز لمجموع مربعات هذا النوع من الانحرافات بالرمز SSR حيث:

$$SSR = \sum\left(\hat{Y} - \bar{Y}\right)^2 = a^2\left[\sum {X_i}^2 - \frac{(\sum X_i)^2}{n}\right]$$

ملاحظة هامة: لا يستخدم القانون الثاني إلا عندما يكون الارتباط **خطيا**.

✓ **الانحرافات غير المفسَّرة**: : وهي الانحرافات بين القيم التقديرية $\hat{Y}$ (التي مصدرها معادلة الانحدار) والقيم الحقيقة Y_i المقابلة لها. وسميت كذلك لأن هذا النوع من الانحرافات لا تفسير له في معادلة الانحدار. يرمز لمجموع مربعات هذا النوع من الانحرافات بالرمز SSE حيث:

$$SSE = \sum\left(Y - \hat{Y}\right)^2$$

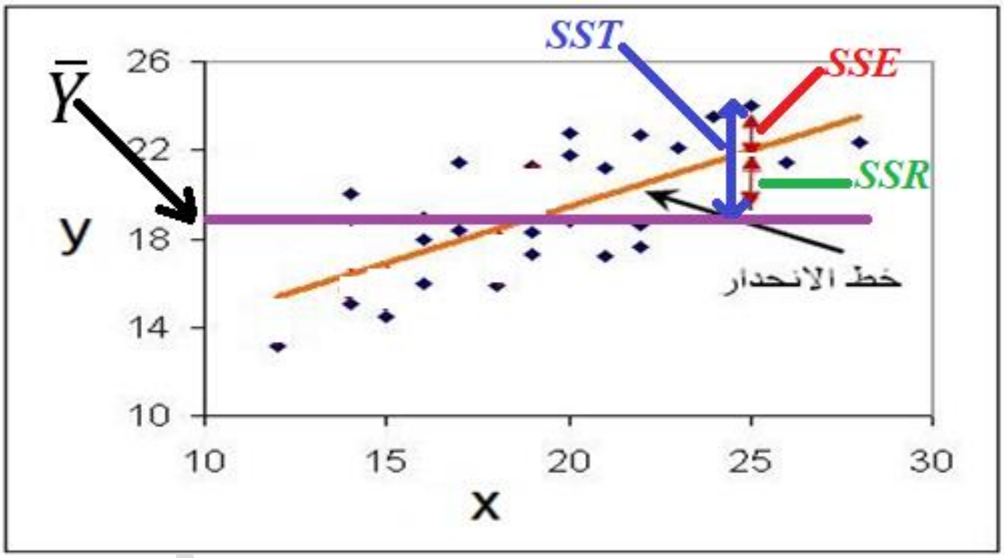
✓ **الانحرافات الكلية**: تنتج عن حاصل جمع النوعين السابقين من الانحرافات، أي أنها عبارة عن انحرافات مفسرة وأخرى غير مفسرة، وهي تجسد الفروق بين القيم الحقيقة Y_i وبين قيمة الوسط الحسابي $\bar{Y}$ (الوسط الحسابي للقيم الحقيقة Y_i). يرمز لها بالرمز SST حيث:

$$SST = SSR + SSE = \sum\left(\hat{Y} - \bar{Y}\right)^2 + \sum\left(Y - \hat{Y}\right)^2 = \sum(Y - \bar{Y})^2$$

$$= \left[\sum Y_i^2 - \frac{(\sum Y_i)^2}{n}\right]$$

ملاحظة هامة: لا يستخدم القانون المبين في السطر الثاني إلا عندما يكون الارتباط **خطيا**.

الشكل رقم 03: الانحرافات الكلية، المفسَّرة وغير المفسَّرة.



المصدر: /https://samehar.wordpress.com/2009/11/11/120909 (بتصرف)

بناءً على ما سبق يمكن النظر لمعامل التحديد باعتباره نسبة "مجموع مربعات الانحرافات المفسَّرة" (SSR) إلى الانحرافات الكلية (SST) حيث:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

وعلى ذلك يمكن حساب معامل بيرسون للارتباط الخطي كما يلي:

$$r_p = \sqrt{R^2} = \sqrt{\frac{SSR}{SST}}$$

ب. معامل ارتباط الرتب: (معامل سبيرمان) r_s

1. **تعريفه**: هو معامل ارتباط ثنائي، يستخدم في حالة المتغيرات الكمية وحتى المتغيرات النوعية القابلة للترتيب، لكنه يعتبر أقل دقة من معامل بيرسون، لكونه يتعامل مع رتب المشاهدات وليس مع المشاهدات نفسها. نلجأ إليه عادة عندما تكون بيانات أحد المتغيرين أو كليهما بيانات نوعية (وصفية) قابلة للترتيب.
2. **حسابه**: يمكن حساب معامل ارتباط الرتب باتباع الخطوات الآتية:
   - ✓ إعطاء رموز رقمية (رتب) لمشاهدات كلٍّ من المتغيرين X و Y.
   - ✓ استخراج الفرق d_i بين رتبتي كل قيمتين متقابلتين من X و Y ووضعه في عمود جديد.
   - ✓ تربيع قيم d_i ووضع ذلك في عمود جديد d_i^2.
   - ✓ حساب معامل "سبيرمان" بتطبيق القانون الآتي: (حيث n هو حجم العينة أو عدد الثنائيات $(x_i\ ,\ y_i)$)

$$r_S = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

ملاحظة هامة: قد نجد بعض المشاهدات متطابقة (سواء في القيمة أو في الصفة) وبالتالي نجد مشكلة في ترتيبها، في هذه الحالة نتبع الخطوات الآتية:

- ➢ نرتب المشاهدات بشكل عادي (كما لو أنها مختلفة ولا تشابه بينها).
- ➢ نحسب الوسط الحسابي لرتب المشاهدات المتشابهة.
- ➢ نستبدل الترتيب القديم لهذه المشاهدات المتشابهة بقيمة هذاالوسط الحسابي.
- ➢ نكمل بقية الخطوات إلى غاية حساب معامل ارتباط الرتب.

3. **خصائصه**: هي ذاتها تقريبا خصائص معامل بيرسون السابق.
4. **مثال توضيحي**: سحبنا عينة من خمسة طلبة، وصدنا تقديراتهم في مقياسي الرياضيات (X) والإحصاء (Y)، ورصدنا ذلك في الجدول الآتي:

الطلبة	1	2	3	4	5
التقدير في الرياضيات X	ممتاز	جيد جدا	جيد	جيد	متوسط
التقدير في الإحصاء Y	جيد جدا	ممتاز	متوسط	جيد	متوسط

المطلوب: حساب معامل ارتباط الرتب بين التقديرين.

الحل: باتباع الخطوات السابقة نحصل على الجدول الآتي:

الطلبة	X	Y	رتب X	رتب Y	d_i	d_i^2
1	ممتاز	جيد جدا	1	2	-1	1
2	جيد جدا	ممتاز	2	1	1	1
3	جيد	متوسط	3.5	4.5	-1	1
4	جيد	جيد	3.5	3	0.5	0.25
5	متوسط	متوسط	5	4.5	0.5	0.25
/	/	/	/	/	**0**	**3.5**

$$r_S = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{6(3.5)}{5(25-1)} = \mathbf{0.82}$$

ومعنى هذا أن هناك علاقة طردية قوية بين ترتيب الطالب في مقياس الرياضيات وترتيبه في مقياس الإحصاء، حيث كلما كان الطالب متفوقا في الأول كان متفوقا أيضا في الثاني. (وهذا منطقي).

ج. معامل الاقتران: r_a

1. **تعريفه**: هناك بعض الحالات نرغب في دراسة العلاقة بين متغيرين نوعيين وغير قابلين للترتيب، فإذا كان هذان المتغيران ثنائيا الإمكانات؛ أي لكل منهما حالتان اثنتان فقط مثل: مدخن وغير مدخن، ذكر وأنثى ... ففي مثل هذه الحالة يمكن الاستعانة بما يسمى "**معامل الاقتران**" r_a .

2. **حسابه**: ولتوضيح كيفية حساب هذا المعامل، لنفرض أن لدينا متغيرا X صفتاه هما: x_1 و x_2، ومتغيرا آخر Y صفتاه هما y_1 و y_2، على النحو الآتي: حيث: الخانات .. n تمثل التكرارات.

X \ Y	y_1	y_2
x_1	n_{11}	n_{12}
x_2	n_{21}	n_{22}

وقد وضع "يول" قانونا لحساب معامل الاقتران كما يلي:

$$r_a = \frac{(n_{11} \times n_{22}) - (n_{12} \times n_{21})}{(n_{11} \times n_{22}) + (n_{12} \times n_{21})}$$

تتراوح قيمة معامل الاقتران بين -1 و +1 .

3. **مثال توضيحي**: قام أحد الباحثين بدراسة العلاقة بين التدخين X والإصابة بأمراض الجهاز التنفسي Y، فجمع بيانات من 335 شخصا، ولخص النتائج في الجدول الآتي:

X \ Y	مصاب	غير مصاب	**المجموع**
مدخن	120	25	**145**
غير مدخن	30	160	**190**
المجموع	**150**	**185**	**335**

$$r_a = \frac{(120 \times 160) - (25 \times 30)}{(120 \times 160) + (25 \times 30)} = \mathbf{0.92}$$

واضح أن هناك اقترانا بين التدخين والغصابة بالامراض التنفسية.

د. معامل التوافق: r_c

1. **تعريفه**: انطلاقا من الحالة السابقة، إذا كان لأحد المتغيرين (أو لكليهما) أكثر من إمكانيتين، هنا نستخدم معامل التوافق، الذي يعتمد -في حسابه على قيمة الإحصاءة "كاي – مربع" χ^2.

2. **حسابه**: يمكن حساب معامل التوافق وفق القانون الآتي:

$$r_c = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

حيث: n هو حجم العينة. وقيمة معامل التوافق دوما موجبة ومحصورة بين 0 و 1.

- ✓ **كيفية استخراج الإحصاءة كاي-مربع**: هناك استخدامات عديدة لاختبار "كاي-مربع"[1]، من بينها أنه يستخدم مقياسا لمدى التفاوت أو التوافق بين ما نتوقعه وبين ما نشاهده فعلا، أي بين التكرارات المتوقعة (النظرية) وبين التكرارات المشاهدة (الحقيقية).يمكن حساب قيمة الإحصاءة بتطبيق القانون الآتي:

[1] سنتطرق إلى هذا الاختبار في السداسي الثاني من السنة الأولى في مقياس الإحصاء 02 (الاحتمالات).

$$\chi^2 = \sum_{i=1}^{K}\sum_{j=1}^{L}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}}$$

حيث:

O_{ij} هو قيمة التكرار الحقيقي (المشاهَد) في الخانة ij.

E_{ij} هو قيمة التكرار المتوقَّع (النظري) في الخانة ij. (طريقة استخراجها مبينة أدناه).

K هو عدد الأسطر R.

L هو عدد الأعمدة C.

ولمزيد من التوضيح نفرض أن لدينا الجدول الآتي الذي يتضمن التكرارت الحقيقية O_{ij} :

Y / X	C_1	C_2	…..	C_L	المجموع
R_1	O_{11}	O_{12}	…..	O_{1L}	$\sum R_1 = \sum_{j=1}^{L} O_{1j}$
R_2	O_{21}	O_{22}	…..	O_{2L}	$\sum R_2 = \sum_{j=1}^{L} O_{2j}$
. . . .					. . . .
R_K	O_{K1}	O_{K2}	…..	O_{KL}	$\sum R_K = \sum_{j=1}^{L} O_{Kj}$
المجموع	$\sum C_1 = \sum_{I=1}^{K} O_{i1}$	$\sum C_2 = \sum_{I=1}^{K} O_{i2}$	…..	$\sum C_L = \sum_{I=1}^{K} O_{iL}$	n

الآن لنحدد التكرارت النظرية E_{ij} في كل خانة من خانات الجدول السابق، يتم ذلك بتطبيق القانون التالي في كل خانة على حدة:

$$E_{ij} = \frac{\sum R_i \times \sum C_j}{n}$$

مثلا بالنسة للخانة الأولى:

$$E_{11} = \frac{\sum R_1 \times \sum C_1}{n}$$

3. **مثال توضيحي**: يبين الجدول الآتي العدد الفعلي للطلبة الذين نجحوا والذين رسبوا (X) في ثلاث مواد دراسية (Y) في قسم علوم التسيير:

Y / X	اقتصاد	تسيير	إحصاء	**المجموع**
ناجح	56	47	50	**153**
راسب	8	14	5	**27**
المجموع	**64**	**61**	**55**	**180**

المطلوب: حدد مقدار التوافق بين النجاح وبين المقاييس الدراسية.

أولا نقوم بحساب قيمة الإحصاءة كاي-مربع، والتي تعتمد على استخراج التكرارات النظرية E_{ij}، التي ندرجها في جدول مستقل (أو في الجدول نفسه).

مثلا الخانة الأولى:

$$E_{11}=\frac{\sum R_1\times\sum C_1}{n}=\frac{153\times 64}{180}=54.40$$

وهكذا مع بقية الخانات، فنحصل على الجدول المقابل:

Y / X	اقتصاد	تسيير	إحصاء	**المجموع**
ناجح	54.40	51.85	46.75	**153**
راسب	9.60	9.15	8.25	**27**
المجموع	**64**	**61**	**55**	**180**

نحسب الآن قيمة الإحصاءة كاي-مربع:

$$\chi^2=\sum_{i=1}^{K}\sum_{j=1}^{L}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}}=\frac{(50-54.4)^2}{54.4}+\frac{(47-51.85)^2}{51.85}+\cdots+\frac{(5-8.25)^2}{8.25}$$

$$=\mathbf{4.85}$$

والآن يمكننا حساب معامل التوافق r_c كما يلي:

$$r_c=\sqrt{\frac{\chi^2}{\chi^2+n}}=\sqrt{\frac{4.85}{4.85+180}}=\mathbf{0.16}$$

ملاحظة هامة:

1. هناك طريقة أخرى لاستخراج قيمة الإحصاءة كاي-مربع دونما حاجة لحساب التكرارات النظرية E_{ij}، وذلك بتطبيق القانون الآتي:

$$\chi^2=n\left[\sum_{i=1}^{K}\sum_{j=1}^{L}\frac{\left(O_{ij}\right)^2}{\sum R_i\times\sum C_j}\right]-n$$

$$=180\left[\frac{56^2}{153\times 64}+\frac{47^2}{153\times 61}+\cdots+\frac{5^2}{27\times 55}\right]-180=\mathbf{4.85}$$

ثم نواصل حساب معامل التوافق كما أوضحنا أعلاه.

2. من المهم الإشارة على أن كلا من معاملي الاقتران والتوافق يعتمدان على ما يسمى "**باختبار الفروض**" أو "**اختبار الفرضيات**"، وذلك للوصول إلى الحكم النهائي حول مدى الاقتران أو التوافق بين الظاهرتين المدروستين، وهو موضوع غير مبرمج في السنة الأولى، لكن سيتم تناوله في برنامج مقياس الإحصاء 03 أو الإحصاء التطبيقي في السنة الثانية. لذلك فإن ما تم تناوله أعلاه بشأن هذين المعاملين يبقى منقوصا إلى حين تناول موضوع "اختبار الفرضيات" لتكتمل الصورة.