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.

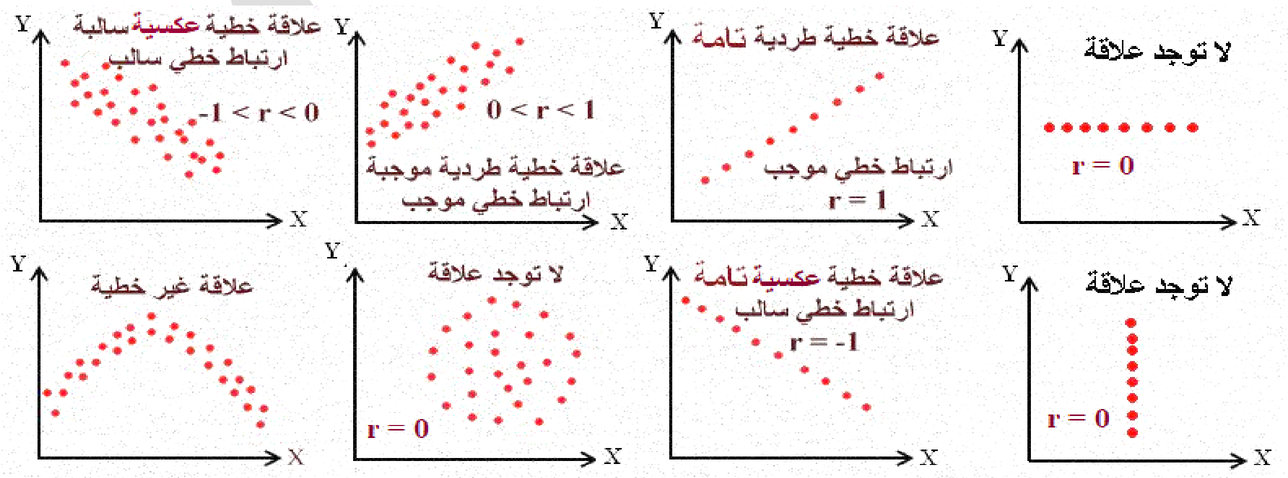
b هو القيمة المقدرة للمتغير Y إذا انعدم المتغير X.

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**."

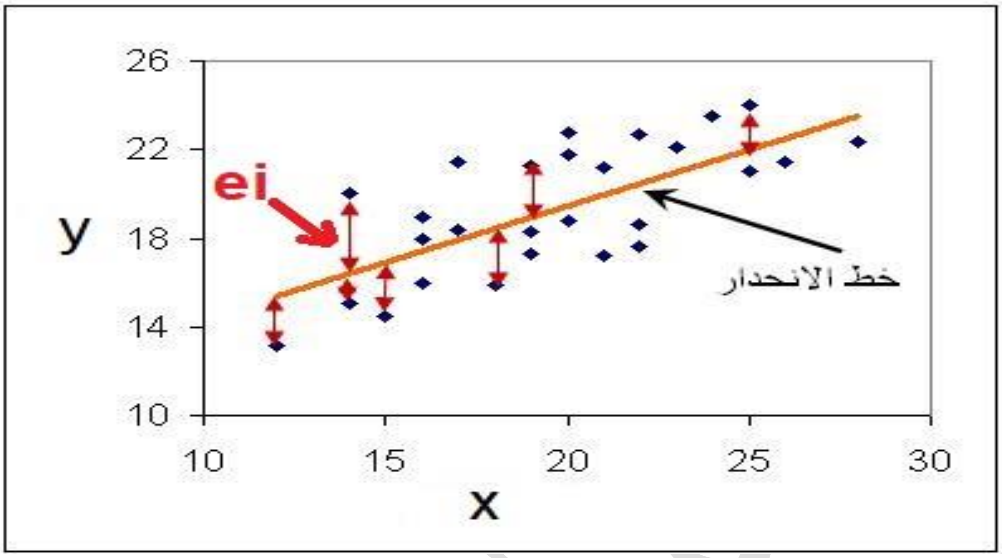
ولأن أفضل خط نبحث عنه يؤدي دورا في سحابة النقاط يشبه دور الوسط الحسابي في مجموعة القيم، فإن مجموع انحرافات القيم الحقيقية Y في السحابة عن القيم المقدرة $\hat{Y}$ الواقعة على الخط يساوي "**الصفر**" حسب الخاصية رقم 01 من خصائص الوسط الحسابي التي رأيناها سابقا.

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

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.



<u>المصدر</u>: 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:

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

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

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

➢ نشتق f بالنسبة للمعلمة b ونضعها تساوي الصفر:

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

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

من (1) و (2) يمكن أن نُكوّن 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: $\boldsymbol{\hat{Y} = 1.45X + 0.75}$

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:

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

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.