تهتم الدراسات الإحصائية بدراسة الارتباط أو الاستقلال الممكن أن يوجد بين متغيرين x و y أو عدة متغيرات لتوزيع احصائي.

لهذا السبب، او لهذا الغرض طورت عدة أدوات ومقاييس لدراسة درجة الارتباط بين متغيرين x و y أو عدة متغيرات، لما تكون هناك خصائص كمية، هذا المقياس يعتمد على مقاييس سببية مثل المتوسطات والتباين، والتي تعد كأساس لتعريف دوال الإنحدار.

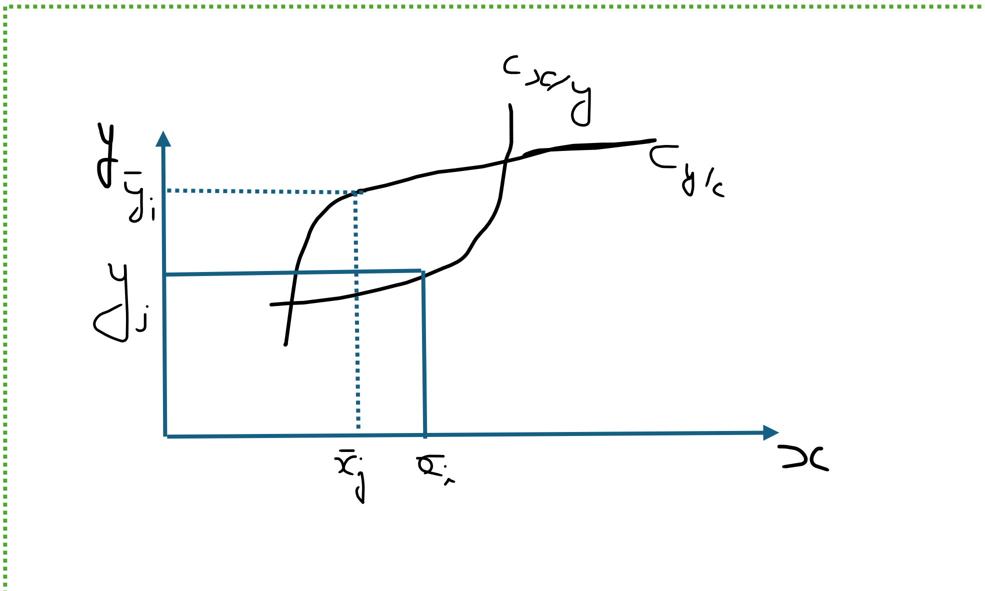
و نميز على العموم، ثلاث أنواع من العلاقات بين متغيرين:

-غياب كلي للعلاقة : لا توجد أي علاقة بين المتغيرين x و y.

-ارتباط كلي : المتغيرين x و y مرتبطين كليا ببعضهما البعض، لدرجة أن معرفة الأول تقتضي بالضرورة معرفة الثاني.

-العلاقة أو الإرتباط النسبي: الإرتباط بين المتغيريين x و y لا تكون 0 % و و لا 100 % ، لكن يمكن أن تؤخذ نسبا بين الحدين ضعيفة، متوسطة، أو قوية.

تعتبر العلاقة نسبية لما يكون التوزيع الإحصائي للثنائية (xi,yi) يشكل سحابة من النقاط متقاربة أو متباعدة.



المنحنى السابق يبين وجود ارتباط بين x و y . لما تكون زاوية انحدار معدومة المنحنيين Cy/x و Cy/xيكونان متطابقان، و يكون هناك ارتباط كلي بين x و y.

و من ناحية أخرى، كلما اقترب المنحنيين Cy/x و Cy/x تنخفض الزاوية المشكلة بين المنحنيين و بالتالي يرتفع أو بقوى الارتباط بين المتغيرين.

ملاحظة: هناك حالات خاصة أن تكون العلاقة غير تبادلية بين المتغيرين و عادة تكون في العلاقات الغير خطية مثل : حالات القطع المكافىء.

1)- y له علاقة ب x ، لاكن x ليس له علاقة ب y.

2)-x له علاقة ب y ، لكن y ليس له علاقة ب x .

و من ما سبق نستنتج شروط دراسة العلاقة بين المتغيرين:

لدراسة العلاقة بين متغيرين يجب توفر الشروط التالية:

-أن تكون بين الظاهرتين المدروستين علاقة جدلية واضحة.

-أن تكون احداهما ناتجة و الأخرى مسببة.

-أن تكون الظاهرتين قابلتين للقباس.

-أن تكون كل القياسات المتحصل عليها متقابلة بالنسبة للزمن أو المكان أو كلاهما معا.

I. التعديل الخطي البسيط:

بافتراضنا وجود مؤشرين (x,y) يمثلان ظاهرة ما y و العنصر الأساسي المؤثر فيها هو x.

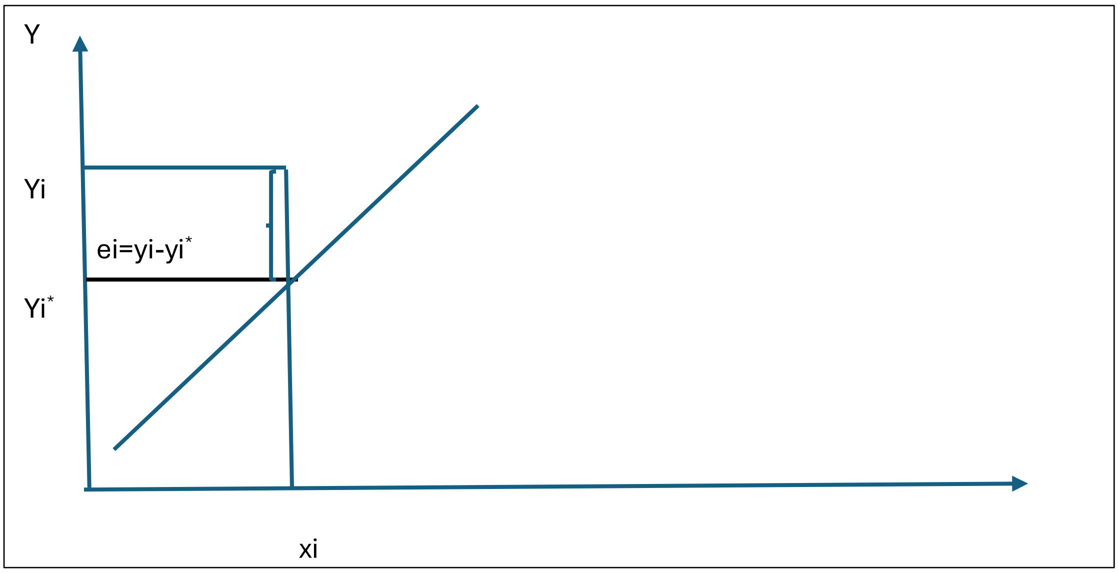
إن العلاقة الخطية البسيطة بين المتغيرين هي العلاقة المستقيمة التي يعبر عنها بواسطة معادلة مستقيم من الشكل $Y=aX+b+e$ و هي تعبر عن أبسط العلاقات الرياضية التي تمثل بها العلاقة بين مؤشرين. من أجل تقدير هذه المعلامات نعتمد على الخطوات التالية:

1-تقدير المعلمات:

سنتطرق لدراسة طريقة المربعات الصغرى:

الهدف من المربعات الصغرى هو إيجاد معادلة التعديل التي تنشأ لتصغير الفروقات بين القيم الملاحظة في الحقيقة و القيم النظرية المعطاة من المنحنى.

لكل قيمة xi يوجد فرق بين القيمة الملاحظة yi و yi^* المقدرة المعرفة انطلاقا من المنحنى، تسمى هذه بالفروقات ei ، و هو معرف في المنحنى بالطريقة التالية:



المنحنى الأمثل الذي يمثل الإتجاه العام أحسن تمثيل هو ذلك الذي يكون فيه الفرق بين كل قيمة $\hat{Y}i$ واقعة عليه و القيمة الحقيقية للظاهرة yi أقل ما يمكن.

في طريقة المربعات الصغرى يوجد نموذج على الطريقة العامة:

$$Yi=\mu i + Ei \qquad (i=1,2,\ldots\ldots\ldots,n)$$

حيث ei هي الفروقات، يجب دائما أن نتأكد أن:

$$E(ei) = 0 \ , VAR (ei) =\sigma^2$$

$$Cov (ei,ej)=0 \ (i\neq j)$$

$$Yi=yi^* +e_i \ (i=1,\ldots\ldots\ldots,n)$$

حيث ei يمكن أن تكون موجبة او سالبة او معدومة .

تكون yi^* القيمة النظرية (لمنحنى التعديل) و يمكن كتابتها بالشكل التالي:

$$Yi^* = f (x)$$

لا يمكن اعتبار هذه المعادلة $Yi^* = f (x)$ كأحسن معادلة لتعديل النقاط (xi,yi) إلا في حالة واحدة، عندما يكون مجموع الفروقات $ei=yi^* -yi$ أصغر إلى أقصى حد و يكتب كالتالي:

$$\boxed{\sum (yi^* - yi)^2 \longrightarrow 0}$$

و تستعمل في حالة النماذج الخطية البسيطة

$$\sum yi^2=\sum yi^2 \ +\sum ei^2$$

و بالتالي يكون لهذا النموذج الخطي أشعة X, E , Y و U ل N مركبة.حيث (N=1,2,.......n).

$$Y= a X +b U+E$$

-تكون المتغيرة الداخلية مفسرة ل Y معرفة.

-تكون المتغيرة الخارجية X معرفة.

-الشعاع U هو شعاع وحدوي (1,.........,1,1,1)U.

-المتغيرةE ، الخطأ، غير معروفة، تمثل الفرق بين القيمة المقدرة و القيمة الحقيقية.

-المعاملات a و b ، التي يكون فيها مجموع المربعات أقل ما يمكن. و بالتالي يكون:

نكتب * أو ^ للتعريف بالقيم المقدرة أي المحسوبة.

$$Y_x^*= a^* \ x +b^*U$$

من مبادئ طريقة المربعات الصغرى اختيار Y^* الأقرب إلى .Y

$$\boxed{\sum (yi^* - yi)^2 \longrightarrow 0min=0}$$

يمر التابع f(a,b) بنهايته الصغرى عند تساوي المشتقتان الجزئيتان.

$$\frac{\delta(a,b)}{\delta b}=0 \ , \ \frac{\delta f[a,b]}{\delta a}=0$$

و تعطى النتائج التالية:

$$\bar{y}=\hat{a}\bar{x}+\hat{b}$$

$$\hat{a}=\frac{\operatorname{cov}(x,y)}{v(x)}$$

$$\hat{a}=\frac{\sum x_i y_i - N\bar{x}\bar{y}}{\sum x_i^2 - N_x^{-2}}$$

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

مع العلم أن:

و بالتالي خط الإنحدار يمر بالنقطة $(\bar{x}, \bar{y})$

*الإنحدارات يمكنها أن تقيس جودة الإنحدار، و بأكثر دقة، تحليل إذا كانت الإستفسارات أكثر أو أقل تفسيرا بالنموذج.

بتفكيك الكمية $(y_i - \bar{y})$ ، التي تمثل انحراف yi بالنسبة لمتوسطها يكون كالتالي:

$$(y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$$

$$(y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + e_i$$

حيث: $(\hat{y}_i - \bar{y})$ تمثل الجزء المفسر عن طريق النموذج و ei هي الجزء من الملاحظات الغير موجودة على خط الانحدار.

إذا وضع التربيع لطرفي المعادلة نتحصل على:

$$\sum(y_i - \bar{y})^2 = \sum(\hat{y}_i - \bar{y})^2 + \sum(y_i - \bar{y}_i)^2$$

$$\boxed{SST \quad = \quad SSR \quad + \quad SSE}$$

SSR : مجموع مربعات الناتجة عن الإنحدارو هي الجزء المفسر من النموذج.

SSE : مجموع مربعات البواقي .

SST : مجموع المربعات الكلي.

هذه المعادلة تكون كأداة لمعرفة مدى أو درجة ملائمة خط الإنحدار ، بواسطة هذه المعادلة يمكننا حساب بعض المعاملات التي يمكنها أن تعطينا مدى تغير yi على $\bar{y}$ SST الذي يفسر عن طريق النموذج SSR. هذا المعامل معرف بمعامل التحديد:

$$\boxed{R^2=SSR/SST=1-(SSE/SST)}$$

معامل التحديد يكون دائما محصور بين 0 و 1، فإذا كان يقترب من الواحد فهذا يعني أن الإنحدار جيد.و هو يقيس مدى جودة النموذج.

2-إختبار الفرضيات:

و بالتالي، يمكننا حساب قيمة تغير البواقي S^2

$$S^2 = 1/(n-2) \,.\, SSE$$

$$=1/(n-2) \sum(yi-b-axi)^2$$

$$= 1/(n-2)\sum(yi-y^*)^2$$

$$=\sum(yi-\bar{y})^2-a^* \sum(xi-x)^2/(n-2)$$

حيث لدينا:

$$\sum(yi-\bar{y})^2=\sum yi^2 -N\, \bar{y}^2$$

$$\sum(xi-\bar{x})^2= \sum \bar{x}i^2-N\, \bar{x}^2$$

أين لدينا n-2 هي درجة الحرية المتعلقة ب SSE و هو يتعلق بعدد المعطيات المستقلة المتاحة لحساب التغير بعد تقدير معلمتي النموذج.

هده المعادلة تكون كأداة لمعرفة مدى أو درجة ملائمة خط الإنحدار، بواسطة هده المعادلة يمكننا حساب بعض المعاملات التي يمكنها أنتعطينا مدى تغير yi على $\bar{Y}$ (SST) الدي يفسر عن طريق النمودج SSR .

I. <u>الإستدلال على المعلمتين في النمودج:</u>

1- إختبار الفرضية المعدومة ل a:

اختبار استدلالية الميل يعتمد على معرفة التأثير الحقيقي للمتغيرة الخارجية x على المتغيرة الداخلية y . الفرضيتين المتضادتين تكتبان كالتالي:

$$\begin{cases} H0 : a=0 \\ H1 : a\neq 0 \end{cases}$$

و يكتب الإختبار كالتالي:

هي تتبع قانون student ب (n-2) كدرجة حرية، و تعتبر منطقة الإختبار لرفض H0 بدرجة خطر α كالتالي:

$$t_a = \left| a^* \right| / S(a)$$

$$S^2(a) = S^2 / \sum (xi - x)^2$$

$$R.C.: \left| ta^* \right| > t(\alpha/2, n-2)$$

في حالة ما كانت القيمة المحسوبة ل t أكبر من القيمة النظرية فهدا يعني أننا سنرفض الفرضية المعدومة لأجل الفرضية المعاكسة ، وعليه يكون التغير في المتغيرة المستقلة يفسر التغير في المتغيرة التابعة بإحتمالية تقدر ب (1-α) . -

و يمكننا ايضا،الإستدلال على أي ثابت آخر حيث يكتب بالطريقة التالية:

$H0 : a = c^t$

$H1 : a \neq c^t$

$$R.C.: \frac{a* - ct}{s(a*)} > t(\alpha/2, n-2)$$

2-مجال الثقة :

a^* معرفة على مجال تعريف a و يمكننا تشكيله أي مجال الثقة عند مستوى ثقة 1-α ، و هو كالتالي:

$$a^* \pm t(\alpha/2, n-2) . s(a)$$

2-اختبار الفرضية المعدومة ل b:

يجب تقدير $s^2(b)$ عن طريق :

$$S^2(b) = \frac{s2\left(\sum xi2\right)}{n.\sum(xi-x)2}$$

$S^2(b)$ يتبع قانون student مع (n-2) درجة حرية:

$$\begin{cases} H0 : b=0 \\ H1 : b\neq 0 \end{cases}$$

$$tcb= \left|b^*\right| /s(b)$$

$$R.C. : \left|tbc\right| > t(\alpha/2, n-2)$$

في حالة ما كانت القيمة المحسوبة ل t أكبر من القيمة النظرية فهدا يعني أننا سنرفض الفرضية المعدومة لأجل الفرضية المعاكسة ، وعليه خط المستقيم لا يمر من المبدأ النقطة (0,0) باحتمالية تقدر ب $1-\alpha$.

-حساب مجال الثقة:

b^* معرفة على مجال تعريف b و يمكننا تشكيله أي مجال الثقة عند مستوى ثقة $1-\alpha$ ، و هو كالتالي:

$$b^* \pm t(\alpha/2, n-2) . s(b)$$

*مستوى الثقة:

احتمالية P ، المعرفة $(1-\alpha)$ ، تسمى مستوى الثقة لمجال الثقة، القيم الأكثر استعمالا ل α هي 0.05 و 0.01، و هي تشكل مالمقدرة مستوى ثقة 0.95 و 0.99 على التوالي.

أيضا إذا اخترنا $\alpha = 0.05$ ، مجال الثقة يتعلق باحتمالية 0.95 لإحتوائه على القيمة الحقيقية θ للمعلمة المقدرة.

طول مجال الثقة هو اذا معيار لعدم الوثوق بالمكانة الحقيقية للقيمة الحقيقية للمعلمة θ المقدرة.

اذا رفعنا من قيمة مجال الثقة طوله سوف يزداد.

II. جدول تحليل التباين ANOVA: تحليل التغير عن طريق الإنحدار الخطي البسيط.

و هو اختبار الإستدلالية الكلي للنموذج. و هو يعتمد على قانون فيشر FISHER عند 1 و n-2 درجات حرية.

H0 : النموذج غير صالح للدراسة.

H1 : النموذج صالح للدراسة.

SOURCE DE VARIANCE	DEGRES DE LIBERTE	SOMME DES CARRES	MOYENNE DES CARRES	Fc
Régression expliquées	1	SSR=SCE= $\sum(y_i^*-\bar{y})^2$	MSR=SSR/1	Fc= MSR/MSE
Résidus	n-2	SSE=SCR= $\sum(y_i^*-y_i)^2$	MSE= SSE/(n-2)	
total	=n-1	SST=SCT= $\sum(y_i-\bar{y})$		

$$F_C > F^{\alpha}_{(1,\ (n-2))}$$

إذا كانت إحصائية فيشر المحسوبة F_C أكبر من إحصائية فيشر النظرية فهذا يعني أن نموذج صالح للدراسة باحتمالية تقدر ب $1-\alpha$. و العكس.

ملاحظة: في الإنحدار الخطي البسيط فقط يمكن أن يكون:

$$Fc = R^2 /((1-R^2)/n-2)$$

$$Fc = \frac{MSR/ddl}{\frac{MSE}{n-2}}$$

$$=\sum(y_i^*-\bar{y})^2 /(\sum e_i^2 /(n-2))$$

$$Fc=t^{*2}$$

و بالتالي هذا الأختبار يمكن أن يكون :

$$H0 : SSR = 0$$

$$H1 : SSR \neq 0$$