المحاضرة الخامسة

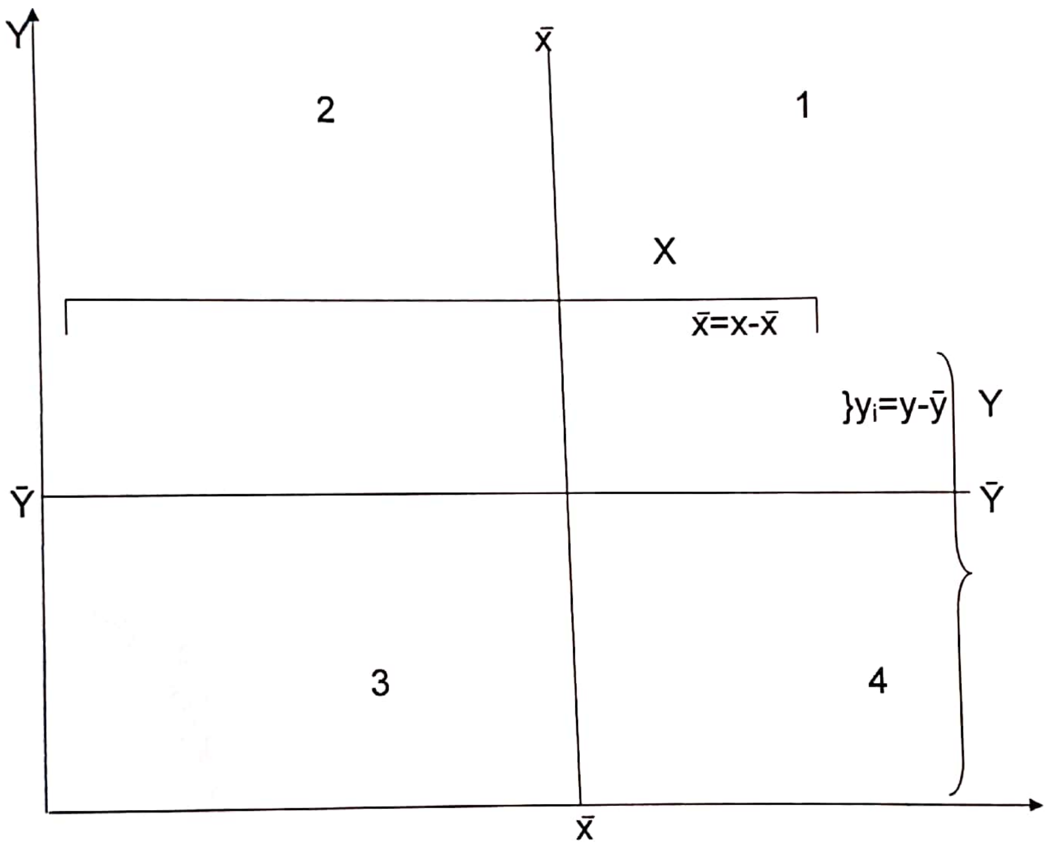
معامل الارتباط

$$t=\frac{(\hat{\beta}-\beta)\sqrt{\Sigma x^2}}{\delta_{\hat{u}}} \cdot \frac{\delta_{\hat{u}}\sqrt{n-2}}{\sqrt{n-2\,\delta_u^2}}$$

$$t=\frac{(\hat{\beta}-\beta)\sqrt{\Sigma x^2}}{\delta_u} \sim t_{n-2}$$

معامل الارتباط

لتفسير الارتباط سوف نعتمد على الشكل الانتشاري. سوف ناخذ عينة من المشاهدات حول x و y و نحاول تمثيلها (كل زوج) في نقطة و الشكل الناتج يكون الشكل الانتشاري.



بتقسيم المستوى بين المحورين x و y الى اربعة اقسام بواسطة مستقيمين متعامدين على المحاور الاحداثية عند النقطتين $\bar{x}$ و $\bar{y}$.

من اجل اية نقطة p ذات احداثيين (x_i,y_i) سوف نعرف الانحرافين

$x_i=x_i-\bar{x}$

$y_i = yi - \bar{y}$

و نلاحظ ما يلي:-

اولا:من اجل جميع النقاط في الربع الاول فان الجداء $x_i y_i$ هو جداء موجب.

ثانيا:من اجل جميع النقاط في الربع الثاني فان الجداء $x_i y_i$ هو جداء سالب.

ثالثا: من اجل جميع النقاط في الربع الثالث فان الجداء $x_i y_i$ هو جداء موجب.

رابعا: من اجل جميع النقاط في الربع الرابع فان الجداء $x_i y_i$ هو جداء سالب.

و اذا اخذنا مجموع جميع الجداءات بالنسبة للنقاط $\sum_{i=1}^{n} x_i y_i$ نلاحظ ان هذا المقدار هو عبارة عن مقياس للترابط بين x و y ذلك اننا ندرك ثلاث حالات.

حـ1: اذا كان الترابط (تناسب طردي) موجبا (بحيث ان اغلب النقاط تتموضع في الربعين الاول و الثالث) فان المجموع $\Sigma x_i y_i > 0$

اي تكون نزعة الى الاشارة الموجبة.

حـ2: اذا كان الترابط تناسبعكسي (سالبا) (بحيث ان اغلب النقاط تتموضع في الربعين الثاني و الرابع) فان المجموع تكون له نزعة الى اشارة سالبة.

حـ3: بينما اذا لم يكون هناك ترابط او علاقة بين x و y فان النقاط سوف تنتشر عبر المربعات الاربعة دون اي قانون. و المجموع $\sum_{i=1}^{n} x_i y_i$ يكون صغيرا جدا و يتناهى الى الصفر.

ملاحظة:الا ان هذا المقياس يعاني من عيبين:-

1- ان قسمته الكمية يمكن ان تزداد بمجرد اضافة مشاهدات اخرى الى العينة.

2- اذا يتاثر ايضا بوحدات القياس بينx و y.

و يتم تصحيح هذين العيبين كما يلي:-

1- فيما يخص العيب الاول يتم تصحيحه بتقسيم المجموع Σxy على حجم العينة فنتحصل بذلك على التباين المشترط للعينة $cov(x,y)=\frac{\Sigma(x,y)}{n}$

2- اما العيب الثاني فيتم تصحيحه بتقسيم المتغيرين x و y على انحرافهما المعياري داخل العينة اي s_y, s_x.

حيث:- $$var(x)=s_x^2=\frac{\Sigma x^2}{n}$$

$$Var(y)=s_y^2=\frac{\Sigma y^2}{n}$$

فيصبح المتغيرين:

$$\frac{x_1}{s_x}, \frac{x_2}{s_x}, -----\frac{x_n}{s_x}$$

$$\frac{y_1}{s_y}, \frac{y_2}{s_y}, -----\frac{y_n}{s_y}$$

$$\frac{\Sigma\frac{x_i}{s_x}\frac{y_i}{s_y}}{n}=\frac{\Sigma x_i y_i}{n s_x s_y}=\frac{cov(x,y)}{\sqrt{var(x)}\sqrt{var(y)}}=\frac{cov(x,y)}{s_x s_y}$$

$$=\frac{\Sigma x_i y_i}{n\sqrt{\frac{\Sigma x^2}{n}}\sqrt{\frac{\Sigma y^2}{n}}}$$

معامل الارتباط البيروسوني $r=\frac{\Sigma x_i y_i}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}}=$

ملاحظة: ايجاد العلاقة بين معامل الارتباط و معامل النحدار

لدينا: $$\beta=\frac{\Sigma}{\Sigma x^2}$$

بضرب كل من البسط و المقام بـ sy

$$\beta=\frac{\Sigma xy}{\Sigma x^2}.\frac{sy}{sy}=\frac{\Sigma xy.sy}{\Sigma x^2.\sqrt{\frac{\Sigma y^2}{n}}}$$

$$=\frac{\Sigma xy.sy}{\sqrt{\Sigma x^2}\sqrt{\Sigma x^2}\sqrt{\frac{\Sigma y^2}{n}}}$$

$$=\frac{\Sigma xy.sy}{\sqrt{\frac{\Sigma x^2}{n}}\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}}$$

$$=\frac{\Sigma xy.sy}{\sqrt{\frac{\Sigma x^2}{n}}\sqrt{\Sigma y^2}.\sqrt{\frac{\Sigma x^2}{n}}}$$

$$\beta=r.\frac{sy}{sx}$$

كما يمكن ايجاد العلاقة بين ميل خط المستقيم و معامل الارتباط

$$r^2=\beta^2\frac{s^2x}{s^2y}=\beta\frac{\Sigma xy}{\Sigma x^2}.\frac{\frac{\Sigma x^2}{n}}{\frac{\Sigma y^2}{n}}$$

و تفيد هذه العلاقة في تحليل التباين $r^2=\beta\frac{\Sigma xy}{\Sigma y^2}$

لدينا من تعريف مستقيم المربعات الصغرى

$$y_i=\hat{y}+e_i \Rightarrow e_i=y_i-\hat{y}_i$$

$$\Sigma y^2=\Sigma\hat{y}_i^2+\Sigma e_i^2+2\Sigma\hat{y}_ie_i$$

$$e_i=y_i-\beta x_i \quad \Longleftarrow \quad 0$$

$$\Sigma\hat{y}e_i=\Sigma\beta x_ie_i$$

$$=\beta[\Sigma x_i(y_i-\beta x_i)]$$

$$=\beta(\Sigma x_iy_i-\beta x_i^2) \longrightarrow 0$$

لان: $\beta=\frac{\Sigma xy}{\Sigma x^2}$

و منه: $\Sigma xy=\beta\Sigma x^2$

و بالتالي: $\Sigma xy-\beta\Sigma x^2=0$

و منه: $\Sigma y_i^2=\Sigma\hat{y}_i^2+\Sigma e_i^2$