Regression Analysis and Correlation Coefficient تحليل الإنحدار ومعامل الإرتباط

التمثيل البياني للإرتباط :

إنّ الخطوة الأولى في دراسة العلاقة بين متغيرين هو إجراء تحليل بياني تصويري حيث يساعد الفحص البصري للبيانات في تزويد المعلومات التالية :

1- تعيين درجة التباين المشترك وهو مؤشر لدرجة الإرتباط بين المتغيرين .

2- تعيين مدى وتوزيع نقاط عينة البيانات .

3- تعيين ظهور حدواث متطرّفة .

4- تعيين شكل العلاقة بين المتغيرين .

5- تعيين نوع العلاقة .

يقصد بالتباين المشترك أنه في حالة وجود تغير في قيمة واحد من المتغيرات سوف تتغير قيمة المتغير الآخر بإسلوب منتظم ، فعلى سبيل المثال يظهر الإزدياد في قيم متغير ما عندما ترتفع قيم المتغير الآخر . لاحظ الشكل التالي :

Linear Regression Analysis: الإنحدار الخطّي البسيط

The least squares method: طريقة المربعات الصغرى

The equation of the line is: $\boxed{\boldsymbol{y = a + bx}}$

Assume the actual observations are: $(y_1, x_1), (y_2, x_2), (y_3, x_3) \dots\ (y_n, x_n)$

Also, let $y =$ actual value, and $\hat{y} =$ calculated value.

So, assume the best fit line is : $\hat{y} = a + bx$

Assume $(y_i - \hat{y}_i) = e_i$ and Summation Square Error (SSE)

$$SSE = \sum {e_i}^2 = \sum (y_i - a - bx)^2$$

To determine the best fit line, follow that:

$$\frac{\partial(SSE)}{\partial a} = -2\sum(y_i - a - bx_i) \qquad \frac{\partial(SSE)}{\partial b} = -2\sum(y_i - a - bx_i)\,.\,x_i$$

Now, let: $\frac{\partial(SSE)}{\partial a} = 0$ and $\frac{\partial(SSE)}{\partial b} = 0$ yields that:

$$\sum y_i = na + b\sum x_i \qquad \dots(1) \qquad \sum y_i x_i = a\sum x_i + b\sum {x_i}^2 \qquad \dots(2)$$

Which are the normal equations وبحل المعادلتين آنياً نحصل على

$$\boxed{\boldsymbol{b = \frac{n\sum y_i x_i - (\sum x_i)(\sum y_i)}{n\sum {x_i}^2 - (\sum x_i)^2}}} \qquad \boxed{\boldsymbol{a = \frac{\sum y_i - b\sum x_i}{n}}}$$

مثال : تمثل القرآت التالية قيم إجهادات القص المناظرة للإجهاد الطبيعي المسلّط على لوح معدني . جد العلاقة الخطّية بين الإجهاد الطبيعي والقص وإحسب قيمة مقاومة القص عند تسليط إجهاد مقداره 10 ميكاباسكال

Strees	x	1	3	4	6	8	9	11	14
Share	y	1	2	4	4	5	7	8	9

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{x^2}$	$\boldsymbol{xy}$	$\boldsymbol{y^2}$
1	1	1	1	1
3	2	9	6	4
4	4	16	16	16
6	4	36	24	16
8	5	64	40	25
9	7	81	63	49
11	8	121	88	64
14	9	196	126	81
$\sum$ 56	40	524	364	256

$$\bar{x} = \frac{56}{8} = 7 \qquad \bar{y} = \frac{40}{8} = 5$$

$$\sum y_i \quad = na \qquad + b\sum x_i$$

$$\sum y_i x_i = a\sum x_i + b\sum {x_i}^2$$

So, $40 = 8a + 56b \qquad \ldots.(1) \times \frac{56}{8} = 7$

So, $364 = 56a + 524b \qquad \ldots.(2)$

So, $280 = 56a + 392b \qquad \ldots.(\bar{1})$

So, $364 = 56a + 524b \qquad \ldots.(2)$

بالطرح

$$280 - 364 = 0 + (392 - 524)b \rightarrow b = \frac{84}{132} = 0.636$$

$$From\ (1) 40 = 8a + 56b \rightarrow 40 = 8a + 56(0.636) \rightarrow a = \frac{4.364}{8} = 0.545$$

Or by laws:

$$b = \frac{n\sum y_i x_i - (\sum x_i)(\sum y_i)}{n\sum {x_i}^2 - (\sum x_i)^2} = \frac{8(364) - (56)(40)}{8(524) - (56)^2} = \frac{7}{11} = 0.636$$

$$a = \frac{\sum y_i - b\sum x_i}{n} = \frac{40 - (0.636)(56)}{8} = \frac{4.364}{8} = 0.545$$

So, $b = 0.636$, and $a = 0.545$

$$\hat{y} = a + bx \rightarrow \hat{y} = 0.545 + 0.636\, x$$

Now, when $x = 10\ MPa$

$$\hat{y} = 0.545 + 0.636\ (10) = 0.545 + 6.36 = 6.905\ MPa$$

Correlation Coefficient: **معامل الإرتباط**

يجري تحليل الإرتباط بعد غشتقاق معادلة تصف العلاقة بين متغيّر ومتغيّر آخر وهذا بسبب إستخدام معامل الإرتباط كمقياس للمطابقة بين معادلة التنبؤ وعينة البيانات المستخدمة لإشتقاق المعادلة .

$$Let\ SSE = \sum e^2 = \sum (\hat{y}_i - y_i)^2$$ Summation Square Error (SSE)

$$Let\ SSR = \sum e^2 = \sum (\hat{y}_i - \bar{y}_i)^2$$ Summation Square Regression (SSR)

$$and\ \ SST = \sum e^2 = \sum (y_i - \bar{y}_i)^2$$ Total Summation Square (SST)

Note that $\boxed{\boldsymbol{SST = SSR + SSE}}$ So,

$$SSR = SST - SSE \Longrightarrow \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = r^2 \qquad \boxed{\boldsymbol{r^2 = \frac{SSR}{SST}}}$$

$$r = \mp\sqrt{\frac{\text{exolaine variation}}{\text{total variation}}} = \sqrt{\frac{SSR}{SST}} = \sqrt{1 - \frac{SSE}{SST}} = \sqrt{\frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}} = \sqrt{1 - \frac{\sum(\hat{y}_i - \bar{y}_i)^2}{\sum(y_i - \bar{y}_i)^2}}$$

مثال: لبيانات المثال السابق إحسب معامل الإرتباط

$$a = 0.545,\ b = 0.636 \text{ and } \hat{y} = 0.545 + 0.636\,x$$

$\boldsymbol{x}$	$\boldsymbol{y}$	$\boldsymbol{\hat{y}}$	$\boldsymbol{e = \hat{y} - y}$	$\boldsymbol{SSE}$ $\boldsymbol{e^2}$	$\boldsymbol{SST}$ $\boldsymbol{(y_i - \bar{y})^2}$	$\boldsymbol{SSR}$ $\boldsymbol{(\hat{y}_i - \bar{y})^2}$
1	1	1.181	-0.181	0.033	16	114.58
3	2	2.453	-0.453	0.205	9	6.49
4	4	3.089	$+0.911$	0.830	1	3.65
6	4	4.361	-0.361	0.130	1	0.41
8	5	5.633	-0.633	0.400	0	0.40
9	7	6.269	$+0.731$	0.534	4	1.61
11	8	7.541	$+0.459$	0.211	9	6.46

14	9	9.449	-0.449	0.202	16	19.79
$\sum = 56$	$\sum = 40$		$\sum \approx 0$	$\sum = 2.545$	$\sum = 56$	$\sum = 53.40$

$$\bar{y} = \frac{40}{8} = 5$$

$$SSE = \sum e^2 = \sum (\hat{y}_i - y_i)^2 = 2.545$$

$$SST = \sum e^2 = \sum (y_i - \bar{y})^2 = 56$$

$$SSR = \sum e^2 = \sum (\hat{y}_i - \bar{y})^2 = 53.40$$

$$r = \mp\sqrt{\frac{SSR}{SST}} = \mp\sqrt{\frac{53.40}{56}} = \mp 0.976 = +97.6\% \quad or$$

$$r = \sqrt{1 - \frac{SSE}{SST}} = \sqrt{1 - \frac{2.545}{56}} = \sqrt{\frac{53.40}{56}} = 97.6\%$$