Esercizio 1

Siano dati i seguenti valori per due variabili:

X	Y
6	4
7	2
8	5
4	2
5	3

1. Si calcolino i coefficienti della retta di regressione usando X come variabile indipendente ed Y come dipendente.
2. Si calcolino i coefficienti della retta di regressione usando Y come variabile indipendente ed X come dipendente.
3. Cambiano i coefficienti nei due casi? Ed il coefficiente di determinazione? Si calcoli il coefficiente di correlazione lineare a partire dal coefficiente di determinazione.
4. Qual è l'unico punto in cui si intersecano le due rette?

Soluzione

1. Si calcolino i coefficienti della retta di regressione usando Y come variabile dipendente ed X come indipendente.

Y	**X**	$x_i \cdot y_i$	x_i^2	$\hat{y}_i$	$(y_i - \bar{y})^2$	$(\hat{y}_i - \bar{y})^2$
4	6	24	36	3.2	0.64	0
2	7	14	49	3.7	1.44	0.25
5	8	40	64	4.2	3.24	1
2	4	8	16	2.2	1.44	1
3	5	15	25	2.7	0.04	0.25
30	16	101	190		6.8	2.5

$\bar{x} = 6$ e $\bar{y} = 3.2$

$\hat{\beta}_1 = \frac{\frac{1}{5}101 - 3.2 \cdot 6}{\frac{1}{5}190 - 6^2} = \frac{1}{2} = 0.5$
$\hat{\beta}_0 = 3.2 - 0.5 \cdot 6 = 0.2$
$R^2 = \frac{2.5}{6.8} = 0.3676$

2. Si calcolino i coefficienti della retta di regressione usando X come variabile dipendente ed Y come dipendente. $\bar{x} = 6$ e $\bar{y} = 3.2$

X	**Y**	$x_i \cdot y_i$	y_i^2	$\hat{x}_i$	$(x_i - \bar{x})^2$	$(\hat{x}_i - \bar{x})^2$
6	4	24	16	6.5882	0	0.3460
7	2	14	4	5.1176	1	0.7786
8	5	40	25	7.3235	4	1.7516
4	2	8	4	5.1176	4	0.7786
5	3	15	9	5.8525	1	0.0218
30	16	101	58		10	3.6766

$\hat{\beta}_1 = \frac{\frac{1}{5}101 - 3.2 \cdot 6}{\frac{1}{5}58 - 3.2^2} = \frac{1}{1.36} = 0.7353$
$\hat{\beta}_0 = 6 - 0.7353 \cdot 3.2 = 3.647$
$R^2 = \frac{3.6766}{10} = 0.36766$

3. Il coefficiente di determinazione non cambia, ma cambiano le equazioni delle due rette. Quindi definire quale variabile usare come dipendente e quale come indipendente è determinante.
In entrambi i casi $\rho = \sqrt{0.3676} = 0.6063$.

4. L'unico punto in cui passano entrambe le rette sarà il baricentro $(6; 3.2)$.

Esercizio 2

Sono stati rilevati su 100 studenti del secondo anno le seguenti variabili Y="numero di esami sostenuti" e X="ore settimanali passate davanti alla televisione". Si è osservato che:

$$\bar{X} = 20 \text{ e } \bar{Y} = 4$$
$$\sigma_X^2 = 60 \text{ e } \sigma_Y^2 = 2$$
$$\frac{1}{n}\sum_i x_i y_i = 68$$

Possiamo prevedere quanti esami mediamente supera uno studente che passa davanti alla televisione 22 ore in una settimana? Se sì, quanti?

Soluzione

$\hat{\beta}_1 = \frac{68-4\cdot 20}{60} = \frac{-12}{60} = -0.2$
$\hat{\beta}_0 = 4 + 0.2 \cdot 20 = 8$
Il modello stimato è $y = 8 - 0.2x \Rightarrow \hat{y} = 8 - 0.2 \cdot 22 = 3.6$.
Sulla base del modello, guardando la TV 22 ore, ci si può aspettare di superare 3.6 esami.