

Esercizi sulla regressione e lo stimatore OLS

Corso di Econometria

Professor Valerio Potì

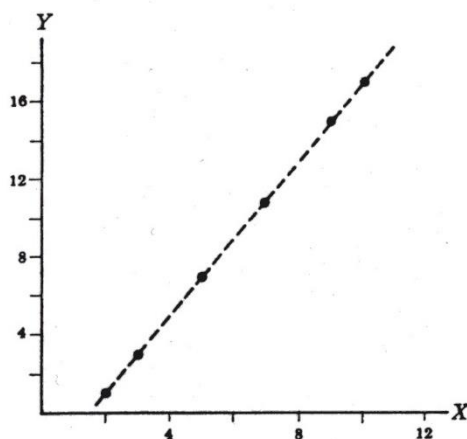
PARTE A

Esercizi, soprattutto di richiamo (alcuni elementari), di fondamenti di analisi matematica della regressione e sulla interpolazione secondo il metodo dei minimi quadrati

A1. Si affrontino i seguenti compiti:

- Costruite una retta che interpoli i dati in tabella.
- Trovare un'equazione per tale retta.

X	2	3	5	7	9	10
Y	1	3	7	11	15	17



A.2 Dal problema precedente, trovare

- Y quando X=4
- Y quando X=15
- Y quando X=0
- X quando Y=7.5
- X quando Y=0.

A.3. Dimostrare che l'equazione della retta passante per i punti (X_1, Y_1) e (X_2, Y_2) è data da

$$Y - Y_1 = \frac{Y_2 - Y_1}{X_2 - X_1} (X - X_1)$$

$$\frac{Y - Y_1}{X - X_1} = \frac{Y_2 - Y_1}{X_2 - X_1}$$

→

$$Y - Y_1 = \frac{Y_2 - Y_1}{X_2 - X_1} (X - X_1)$$

→

$$Y = Y_1 + \frac{Y_2 - Y_1}{X_2 - X_1} (X - X_1)$$

$$Y = Y_1 + \frac{Y_2 - Y_1}{X_2 - X_1} X - \frac{Y_2 - Y_1}{X_2 - X_1} X_1$$

$$Y = Y_1 - \frac{Y_2 - Y_1}{X_2 - X_1} X_1 + \frac{Y_2 - Y_1}{X_2 - X_1} X$$

E dunque,

$$Y = a_0 + a_1 X$$

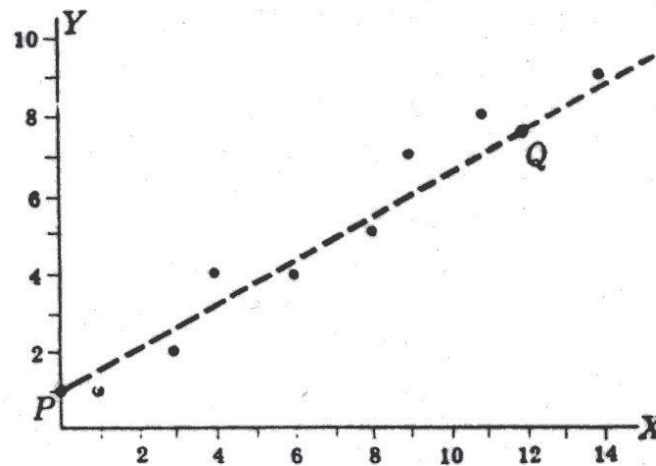
dove

$$a_0 = Y_1 - \frac{Y_2 - Y_1}{X_2 - X_1} X_1 = Y_1 - a_1 X_1$$

$$a_1 = \frac{Y_2 - Y_1}{X_2 - X_1}$$

- A.4. Trovate (a) il coefficiente angolare, (b) l'equazione, (c) l'ordinata all'origine e (d) l'intercetta sull'asse delle X della retta passante per i punti (1,5) e (4,-1).
- A.5. Si affrontino i seguenti compiti:
 a) Costruite una retta che interpoli i dati in tabella.
 b) Trovate l'equazione di tale retta.

X	1	3	4	6	8	9	11	14
Y	1	2	4	4	5	7	8	9



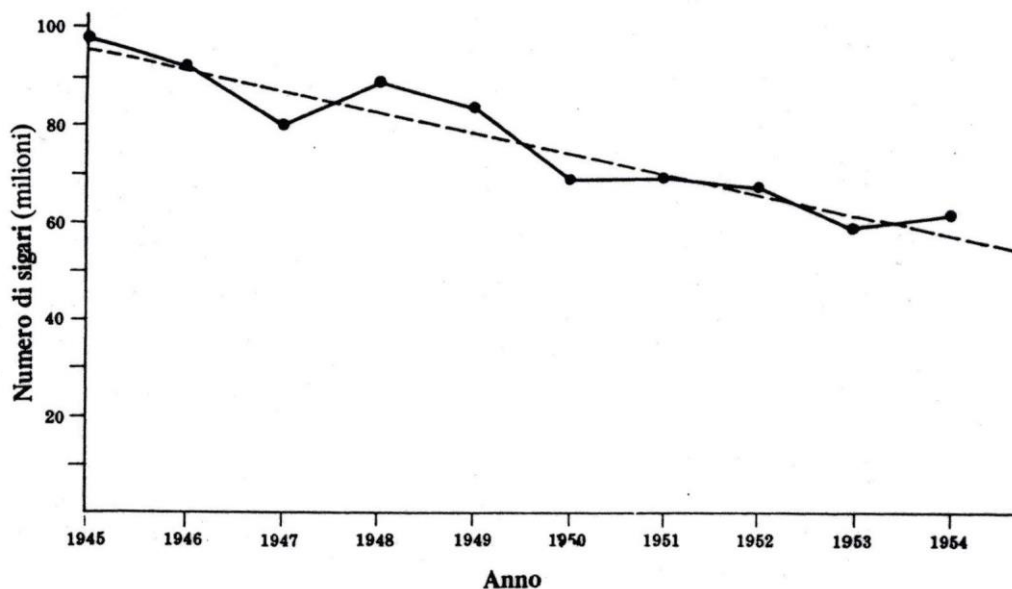
- A.6. Interpolate i dati del problema precedente con una retta dei minimi quadrati, usando X come variabile indipendente.
- A.7. In tabella è riportata la produzione di acciaio di un certo paese in milioni di tonnellate durante gli anni 1946-1956.

Anno	Produzione di acciaio (milioni di tonnellate)
1946	66,6
1947	84,9
1948	88,6
1949	78,0
1950	96,8
1951	105,2
1952	93,2
1953	111,6
1954	88,3
1955	117,0
1956	115,2

- Tracciate il grafico dei dati
- Trovate l'equazione della retta dei minimi quadrati interpolante i dati.
- Stimate la produzione di acciaio per gli anni 1957 e 1958 e confrontate con i veri valori, che sono rispettivamente 112,7 e 85,3 milioni di tonnellate.
- Stimate la produzione di acciaio per gli anni 1945 e 1944 e confrontate con i veri valori, che sono 79,7 e 89,6 milioni di tonnellate.

A.8. La tabella dà la produzione di sigari negli Stati Uniti per gli anni 1945 – 1954.

Anno	1945	1946	1947	1948	1949	1950	1951	1952	1953	1954
Numero di sigari (milioni)	98,2	92,3	80,0	89,1	83,5	68,9	69,2	67,1	58,3	61,2



Trovate l'equazione della retta dei minimi quadrati interpolante i dati.

A.9. La tabella seguente riporta le rispettive masse X ed Y di un campione di 12 padri e dei loro figli primogeniti:

Osservazione	Padre	Figlio
1	62	66

2	63	66
3	64	65
4	65	68
5	66	65
6	67	67
7	67	68
8	68	69
9	68	71
10	69	68
11	70	68
12	71	70

- a) Costruite un diagramma di dispersione
 - b) Trovate la retta di regressione dei minimi quadrati di Y su X
- A.10. Calcolate, per i dati del problema precedente, quanto segue:
- a) la devianza totale
 - b) la devianza residua
 - c) la devianza spiegata
- A.11 Trovate (a) il coefficiente di determinazione e (b) il coefficiente di correlazione per i dati per problema precedente.

Soluzioni

Esercizio A.1

- a) Si riportano i punti (2,1), (3,3), (5,7), (7,11), (9,15) e (10,17) su un sistema di coordinate cartesiane.

Tutti i punti giacciono su una retta, che per conseguenza interpola i dati esattamente.

- b) Vogliamo determinare l'equazione della retta data da

$$Y = a_0 + a_1 X$$

Sono quindi necessari solo 2 punti. Scegliamo i punti (2,1) e (3,3).

Per il punto (2,1), $X=2$ e $Y=1$.

Avremo quindi

$$1 = a_0 + 2a_1.$$

Analogamente per il punto (3,3), $X=3$ e $Y=3$. Sostituendo,

$$3 = a_0 + 3a_1.$$

Risolvendo simultaneamente,

$$a_0 = -3$$

$$a_1 = 2,$$

quindi

$$Y = -3 + 2X$$

$$Y = 2X - 3.$$

Esercizio A.2

L'enunciato dell'esercizio essenzialmente richiede che assumiamo che la stessa relazione $Y = 2X - 3$ appena trovata valga anche per valori di X e Y diversi da quelli utilizzati per trovarla.

- a) Se $X = 4$, $Y = 8 - 3 = 5$. Dato che vogliamo trovare il valore di Y corrispondente a un valore di X compreso fra due valori dati di X , questo procedimento è detto *interpolazione lineare*.
- b) Se $X = 15$, $Y = 2(15) - 3 = 27$. Dato che vogliamo trovare il valore di Y corrispondente ad un valore di X estremo ai valori dati di X , tale procedimento è detto *estrapolazione lineare*.
- c) Se $X = 0$, $Y = 2(0) - 3 = 0 - 3 = -3$.
Il valore di Y quando $X = 0$ è detto *intercetta* sull'asse delle Y . È il valore di Y nel punto in cui la retta (eventualmente prolungata) taglia l'asse delle Y .
- d) Se $Y = 7.5$, $2X = 7.5 + 3 = 10.5$, $X = 10.5/2 = 5.25$.
- e) Se $Y = 0$, $0 = 2X - 3$: quindi $2X = 3$ e $X = 1.5$. Il valore di X quando $Y = 0$ è detto intercetta sull'asse delle Y .
- f) È il valore di X nel punto in cui la retta (eventualmente prolungata) taglia l'asse delle X .

Esercizio A.3

L'equazione della retta è

$$Y = a_0 + a_1 X$$

Poichè (X_1, Y_1) giace sulla retta, abbiamo

$$Y_1 = a_0 + a_1 X_1.$$

Poichè (X_2, Y_2) giace sulla retta,

$$Y_2 = a_0 + a_1 X_2.$$

Sottraiamo le prime due equazioni e otteniamo

$$Y - Y_1 = a_1 (X - X_1)$$

Sottraendo la seconda e la terza equazione e otteniamo

$$Y_2 - Y_1 = a_1 (X_2 - X_1),$$

quindi

$$a_1 = \frac{Y_2 - Y_1}{X_2 - X_1}$$

da cui l'equazione desiderata:

$$Y - Y_1 = a_1 (X - X_1) = \frac{Y_2 - Y_1}{X_2 - X_1} (X - X_1)$$

L'espressione $a_1 = \frac{Y_2 - Y_1}{X_2 - X_1}$ è il coefficiente angolare della retta per cui a volte si usa il simbolo m .

Esercizio A.4

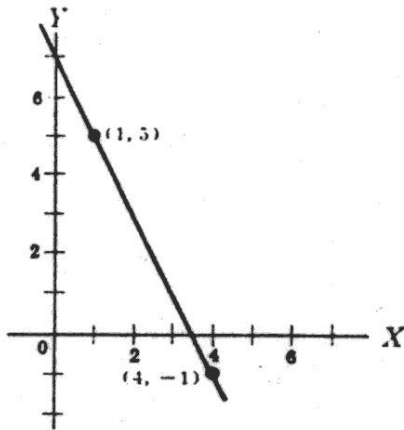
(a) Utilizzando l'equazione

$$Y - Y_1 = \frac{Y_2 - Y_1}{X_2 - X_1} (X - X_1)$$

ovvero le equazioni per i coefficienti dell'equazione della retta derivate dalla stessa equazione, si ha

$$m = \frac{Y_2 - Y_1}{X_2 - X_1} = \frac{-1 - 5}{4 - 1} = -2$$

Il segno negativo del coefficiente angolare indica che, al crescere di X , Y diminuisce:



(b) L'equazione della retta è

$$Y - Y_1 = m (X - X_1)$$

ovvero

$$Y - 5 = -2(X - 1)$$

ossia

$$Y = 7 - 2X.$$

(c) L'ordinata all'origine, che è il valore di Y quando X=0, è data da $Y = 7 - 2(0) = 7$, come d'altra parte si vede dal grafico.

(d) L'intercetta sull'asse delle X è il valore di X quando Y=0. Sostituendo Y=0 nell'equazione $Y = 7 - 2X$ abbiamo

$$0 = 7 - 2X \text{ cioè}$$

$$2X = 7 \text{ ossia } X = 3.5 .$$

Questo valore può anche essere visto dal grafico.

Esercizio A.5

a) Riportiamo i vari punti (1,1), (3,2), ecc. su un sistema di coordinate cartesiane come in figura.

La retta interpolante è tracciata empiricamente nella figura. Per tracciarla nel modo corretto ci si dovrà servire del metodo dei minimi quadrati.

b) Per ottenere l'equazione della retta, scegliamo due punti qualsiasi sulla retta, come ad esempio P e Q, che valgono approssimativamente (0,1) e (12, 7.5).

L'equazione della retta è

$$Y = a_0 + a_1 X,$$

che applicata ai punti dati ci da

$$1 = a_0 + a_1 (0)$$

$$7.5 = a_0 + 12 a_1$$

Si ottiene quindi

$$a_0 = 1$$

$$a_1 = 6.5/12 = 0.542.$$

Si ottiene così

$$Y = 1 + 0.542 X.$$

Altro metodo, usando

$$Y - Y_1 = \frac{Y_2 - Y_1}{X_2 - X_1} (X - X_1)$$

$$Y - 1 = [(7.5 - 1)/(12 - 0)] (X - 0)$$

$$Y - 1 = 0.542 X$$

$$Y = 1 + 0.542 X$$

Esercizio A.6

L'equazione della retta è

$$Y = a_0 + a_1 X + u$$

Le equazioni normali, dove N denota il numero di osservazioni, sono

$$\sum Y = a_0 N + a_1 \sum X$$

$$\sum (XY) = a_0 \sum X + a_1 \sum X^2$$

Il calcolo delle somme viene ordinato per colonne:

X	Y	X ²	XY	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
$\Sigma X = 56$	$\Sigma Y = 40$	$\Sigma X^2 = 524$	$\Sigma XY = 364$	$\Sigma Y^2 = 256$

Quindi abbiamo:

$$40 = a_0 \cdot 8 + a_1 \cdot 56$$

$$364 = a_0 \cdot 56 + a_1 \cdot 524$$

Dato che ci sono N = 8 paia di valori di X e Y, N=8, e le equazioni normali diventano

$$\begin{aligned} 8 a_0 + 56 a_1 &= 40 \\ 56 a_0 + 524 a_1 &= 364 \end{aligned}$$

Risolvendo simultaneamente,

$$\begin{aligned} a_0 + 56/8 a_1 &= 40/8 \Rightarrow a_0 + 7 a_1 = 5 \Rightarrow a_0 = 5 - 7 a_1 \\ 56 a_0 + 524 a_1 &= 364 \Rightarrow 56 (5 - 7 a_1) + 524 a_1 = 364 \\ \Rightarrow 280 - 392 a_1 + 524 a_1 &= 364 \\ \Rightarrow -392 a_1 + 524 a_1 &= 84 \\ \Rightarrow 132 a_1 &= 84 \\ \Rightarrow a_1 &= 84 / 132 = 0.63 \\ \Rightarrow a_0 &= 5 - 7 \times 0.63 \\ \Rightarrow a_0 &= 0.55 \end{aligned}$$

$$\begin{aligned} 8a_0 + 56a_1 &= 40 \\ a_0 + 56/8a_1 &= 40/8 \\ a_0 + 7a_1 &= 5 \\ a_0 &= 5 - 7a_1 \\ 524a_1 &= -56a_0 + 364 \\ a_1 &= -\frac{56}{524}a_0 + \frac{364}{524} \\ a_1 &= -0.106870229 a_0 + 0.694656489 \\ a_0 &= 5 - 7(-0.106870229 a_0 + 0.694656489) \\ a_0 &= 5 + 0.748a_0 - 4.863 \\ a_0 - 0.748a_0 &= 5 - 4.863 \\ 0.25a_0 &= 0.14 \\ a_0 &= \frac{0.14}{0.25} = 0.545 \\ a_1 &= -\frac{56}{524}0.545 + \frac{364}{524} = 0.636 \end{aligned}$$

Ovvero, utilizzando le formule del compendio,

$$\begin{aligned} E(X) &= 56/8 = 7 \\ E(Y) &= 40/8 = 5 \\ \text{Cov}(X, Y) &= E(XY) - E(X) E(Y) = 364/8 - (56/8)(40/8) = 10.5 \\ \text{Var}(X) &= E(X^2) - E(X)^2 = 524/8 - (56/8)^2 = 16.5 \end{aligned}$$

e quindi

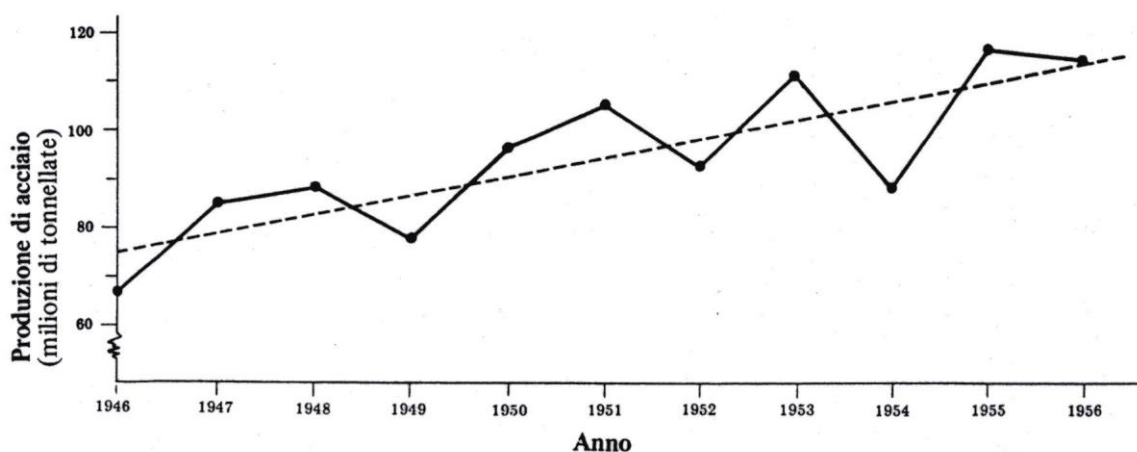
$$\begin{aligned} a_1 &= 10.5/16.5 = 0.63. \\ a_0 &= 5 - 0.63 \times 7 = 0.55 \end{aligned}$$

In entrambi i modi, le soluzioni sono identiche. La retta dei minimi quadrati richiesta è dunque

$$Y = 0.545 + 0.636 X$$

Esercizio A.7 (Applicazione alle serie temporali)

a) Abbiamo:



b) La soluzione delle equazioni normali è

$$\begin{aligned} a_0 &= \frac{\sum Y}{N} - a_1 \frac{\sum X}{N} \\ &= \bar{Y} - a_1 \bar{X} \end{aligned}$$

$$a_1 = (\sum XY - a_0 \sum X) / \sum X^2$$

È chiaro che questa soluzione si semplifica molto se X e Y hanno media nulla. Lavoriamo dunque con

$$x = X - \bar{X} \quad e \quad y = Y - \bar{Y}.$$

La soluzione allora diventa

$$\begin{aligned} a_0 &= \bar{y} - a_1 \bar{x} = 0 \\ a_1 &= (\sum xy - a_0 \sum x) / \sum x^2 \\ &= (\sum xy) / \sum x^2 \end{aligned}$$

I calcoli necessari possono essere organizzati come in tabella:

Anno	X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	x^2	xy
1946	0	66,6	-5	-28,4	25	142,0
1947	1	84,9	-4	-10,1	16	40,4
1948	2	88,6	-3	-6,4	9	19,2
1949	3	78,0	-2	-17,0	4	34,0
1950	4	96,8	-1	1,8	1	-1,8
1951	5	105,2	0	10,2	0	0
1952	6	93,2	1	-1,8	1	-1,8
1953	7	111,6	2	16,6	4	33,2
1954	8	88,3	3	-6,7	9	-20,1
1955	9	117,0	4	22,0	16	88,0
1956	10	115,2	5	20,2	25	101,0
	$\sum X = 55$ $\bar{X} = 5$	$\sum Y = 1045,4$ $\bar{Y} = 95,0$			$\sum x^2 = 110$	$\sum xy = 434,1$

L'equazione stimata richiesta diventa

$$\hat{y} = (434,1/110) x$$

ossia

$$\hat{y} = 3,95 x$$

Che può essere riscritta nella forma

$$\begin{aligned} \hat{Y} - 95,0 &= 3,95 (X - 5) \quad \text{oppure} \\ \hat{Y} &= 75,2 + 3,95 X \end{aligned}$$

Dove l'origine $X = 0$ è l'anno 1946 e l'unità di misura di X è l'anno.

- c) Usiamo l'equazione del trend $Y = 75,2 + 3,95 X$, dove $X=0$ corrisponde al 1951. Quindi gli anni 1957 e 1958 corrispondono rispettivamente a $X=6$ e $X=7$.

Per $X=6$, $Y = 75,2 + 3,95 (6) = 98,9$, che non è molto prossimo al valore reale 112,7.

Per $X=7$, $Y = 75,2 + 3,95 (7) = 102,85$, che non è molto prossimo al valore reale 85,3.

Ciò illustra il rischio implicato dal procedimento dell'estrapolazione.

Gli stessi risultati possono essere ottenuti usando l'equazione del trend $Y = 75.2 + 3.95X$ che ha come origine l'anno 1946, ponendo rispettivamente $X = 11$ e $X = 12$.

d) Usando la retta del trend $Y = 75.2 + 3.95X$ con $X = -1$ e $X = -2$, troviamo i valori

$$Y = 75.2 + 3.95(-1) = 71.2$$

e

$$Y = 75.2 + 3.95(-2) = 67.3$$

Esercizio A.8

Compiliamo le colonne delle somme:

Anno	X	Y	$x = X - \bar{X}$	$y = Y - \bar{Y}$	x^2	xy
1945	0	98,2	-4,5	21,4	20,25	-96,30
1946	1	92,3	-3,5	15,5	12,25	-54,25
1947	2	80,0	-2,5	3,2	6,25	-8,00
1948	3	89,1	-1,5	12,3	2,25	-18,45
1949	4	83,5	-0,5	6,7	0,25	-3,35
1950	5	68,9	0,5	-7,9	0,25	-3,95
1951	6	69,2	1,5	-7,6	2,25	-11,40
1952	7	67,1	2,5	-9,7	6,25	-24,25
1953	8	58,3	3,5	-18,5	12,25	-64,75
1954	9	61,2	4,5	-15,6	20,25	-70,20
	$\Sigma X = 45$ $\bar{X} = 4,5$	$\Sigma Y = 767,8$ $\bar{Y} = 76,8$			$\Sigma x^2 = 82,5$	$\Sigma xy = -354,9$

L'equazione richiesta $\hat{y} = \left(\frac{\Sigma xy}{\Sigma x^2}\right) x$ diventa

$$\hat{y} = (-354.9/82.5) x \text{ ossia } \hat{y} = -4.30 x$$

Che può essere scritta nella forma Y

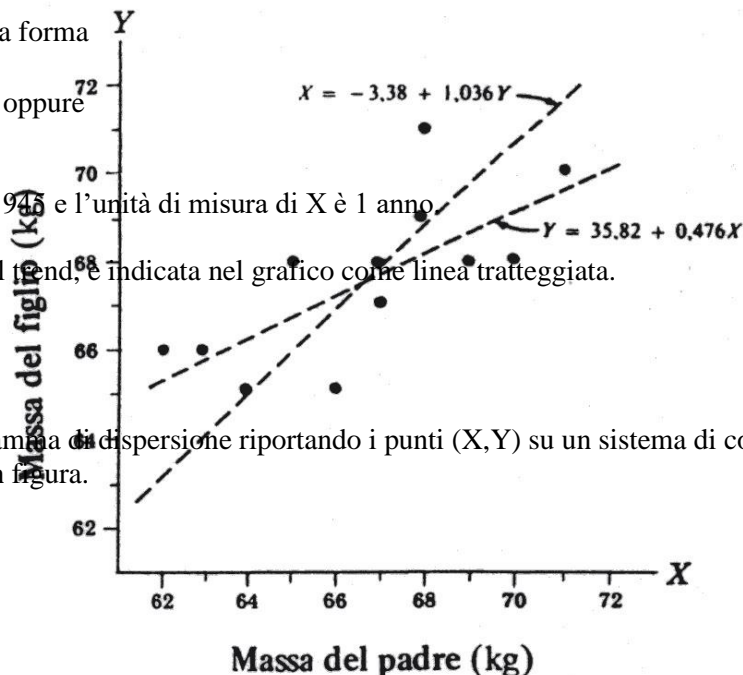
$$\hat{Y} - 76.8 = -4.30 (X - 4.5) \text{ oppure } Y = 96.2 - 4.30 X$$

Dove l'origine $X = 0$ è il 1945 e l'unità di misura di X è 1 anno

Questa retta, detta retta del trend, è indicata nel grafico come linea tratteggiata.

Esercizio A.9

a) Si ottiene il diagramma di dispersione riportando i punti (X,Y) su un sistema di coordinate cartesiane come in figura.



- b) La retta di regressione di Y su X è data da $Y = a_0 + a_1 X$, dove a_0 e a_1 vengono ottenute risolvendo le equazioni normali

$$\begin{aligned}\Sigma Y &= a_0 N + a_1 \Sigma X \\ \Sigma XY &= a_0 \Sigma X + a_1 \Sigma X^2\end{aligned}$$

Lavorando prima con X e Y a cui sia stata preventivamente sottratta la rispettiva media campionaria e calcolando le dovute somme, si ottiene

$$\begin{aligned}0 &= a_0 \cdot 12 + a_1 \cdot 0 \\ 40.33 &= a_0 \cdot 0 + a_1 \cdot 84.67\end{aligned}$$

e dunque

$$\begin{aligned}0 &= 12 a_0 \Rightarrow a_0 = 0 \\ 40.33 &= a_1 \cdot 84.67 \Rightarrow a_1 = 40.33/84.67 = 0.476\end{aligned}$$

Lavorando ora con le X e Y originarie (a cui non sia stata preventivamente sottratta la rispettiva media campionaria) e calcolando le dovute somme, si ottiene

$$\begin{aligned}811 &= a_0 \cdot 12 + a_1 \cdot 800 \\ 54107 &= a_0 \cdot 800 + a_1 \cdot 53418\end{aligned}$$

e dunque

$$\begin{aligned}12 a_0 + 800 a_1 &= 811 \\ 800 a_0 + 53418 a_1 &= 54417\end{aligned}$$

da cui

$$\begin{aligned}a_0 &= 35.82 \\ a_1 &= 0.476\end{aligned}$$

e dunque

$$Y = 35.82 + 0.476 X$$

Esercizio A.10

Calcolando le somme sottoindicate si verifica che:

- a) Devianza totale = $\sum(Y - \bar{Y})^2 = \sum y^2 = 38.92$
- b) Devianza residua = $\sum(Y - Y_{stim})^2 = 19.70$
- c) Devianza spiegata = $\sum(Y_{stim} - \bar{Y})^2 = 38.92 - 19.70 = 19.22$.

Per i calcoli possiamo usare la tabella seguente:

Osservazione	Padre	Figlio
1	62	66
2	63	66
3	64	65
4	65	68
5	66	65
6	67	67
7	67	68
8	68	69
9	68	71
10	69	68
11	70	68
12	71	70

Esercizio A.11

- a) Coefficiente di determinazione = $R^2 = \text{devianza spiegata} / \text{devianza totale} = 19.22 / 38.92 = 0.4938$
- b) Coefficiente di correlazione = $r = \pm\sqrt{0.4938} = \pm 0.7027$.

Dato che la variabile Y_{stim} aumenta all'aumentare di X, la correlazione è positiva e scriviamo quindi $r = 0.7027$.