

MATEMATICA PER L'ANALISI DEI DATI

Capitolo 1: Statistica Descrittiva

1.3 Statistica Descrittiva Bivariata

Giovanni Amendola

Corso di laurea triennale in Informatica
Università della Calabria

Anno Accademico 2020/2021

Diagramma di dispersione

Definizione (Diagramma di dispersione o scatterplot)

Date due variabili statistiche su uno stesso campione, il **diagramma di dispersione** o **scatterplot** è un grafico che rappresenta le coppie di osservazioni come punti su un piano cartesiano.

Diagramma di dispersione

Definizione (Diagramma di dispersione o scatterplot)

Date due variabili statistiche su uno stesso campione, il **diagramma di dispersione** o **scatterplot** è un grafico che rappresenta le coppie di osservazioni come punti su un piano cartesiano.

Esempio (Dall'Esempio 23 del Bramanti)

Variabili statistiche: "percentuale di famiglie sotto la soglia di povertà" e "numero di vittime di omicidio in un dato periodo storico" osservate sullo stesso campione di 10 città degli Stati Uniti.

<i>povertà</i>	<i>omicidi</i>
23.60	2707
13.80	1137
18.70	1451
6.10	848
6.20	315
8.10	1130
5.50	412
10.20	732
10.10	1893
6.00	538

Diagramma di dispersione

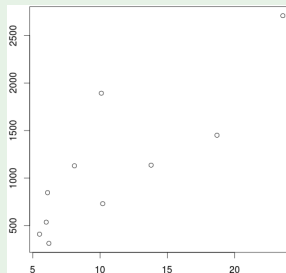
Definizione (Diagramma di dispersione o scatterplot)

Date due variabili statistiche su uno stesso campione, il **diagramma di dispersione** o **scatterplot** è un grafico che rappresenta le coppie di osservazioni come punti su un piano cartesiano.

Esempio (Dall'Esempio 23 del Bramanti)

Variabili statistiche: "percentuale di famiglie sotto la soglia di povertà" e "numero di vittime di omicidio in un dato periodo storico" osservate sullo stesso campione di 10 città degli Stati Uniti.

povertà	omicidi
23.60	2707
13.80	1137
18.70	1451
6.10	848
6.20	315
8.10	1130
5.50	412
10.20	732
10.10	1893
6.00	538



Covarianza

Uno scatterplot è una valutazione **qualitativa** della correlazione. Vogliamo una misura **quantitativa**. Come stabilire se le variabili statistiche sono correlate?

Covarianza

Uno scatterplot è una valutazione **qualitativa** della correlazione. Vogliamo una misura **quantitativa**. Come stabilire se le variabili statistiche sono correlate?

- Se x_i è a sinistra di $\bar{x}$ e y_i a sinistra di $\bar{y}$ (o entrambi a destra), allora il prodotto delle due differenze sarà positivo → **correlazione diretta**.

Covarianza

Uno scatterplot è una valutazione **qualitativa** della correlazione. Vogliamo una misura **quantitativa**. Come stabilire se le variabili statistiche sono correlate?

- Se x_i è a sinistra di $\bar{x}$ e y_i a sinistra di $\bar{y}$ (o entrambi a destra), allora il prodotto delle due differenze sarà positivo → **correlazione diretta**.
- Se x_i è a sinistra di $\bar{x}$ e y_i a destra di $\bar{y}$ (o viceversa), allora il prodotto delle due differenze sarà negativo → **correlazione inversa**.

Covarianza

Uno scatterplot è una valutazione **qualitativa** della correlazione. Vogliamo una misura **quantitativa**. Come stabilire se le variabili statistiche sono correlate?

- Se x_i è a sinistra di $\bar{x}$ e y_i a sinistra di $\bar{y}$ (o entrambi a destra), allora il prodotto delle due differenze sarà positivo → **correlazione diretta**.
- Se x_i è a sinistra di $\bar{x}$ e y_i a destra di $\bar{y}$ (o viceversa), allora il prodotto delle due differenze sarà negativo → **correlazione inversa**.

Definizione (Covarianza)

Sia $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ una successione di n osservazioni congiunte di due variabili. Si dice covarianza delle due variabili x e y la quantità

$$\sigma_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

Covarianza

Uno scatterplot è una valutazione **qualitativa** della correlazione. Vogliamo una misura **quantitativa**. Come stabilire se le variabili statistiche sono correlate?

- Se x_i è a sinistra di $\bar{x}$ e y_i a sinistra di $\bar{y}$ (o entrambi a destra), allora il prodotto delle due differenze sarà positivo → **correlazione diretta**.
- Se x_i è a sinistra di $\bar{x}$ e y_i a destra di $\bar{y}$ (o viceversa), allora il prodotto delle due differenze sarà negativo → **correlazione inversa**.

Definizione (Covarianza)

Sia $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ una successione di n osservazioni congiunte di due variabili. Si dice covarianza delle due variabili x e y la quantità

$$\sigma_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

Proposizione (dimostrarla per esercizio)

$$\sigma_{xy} = \left(\frac{1}{n} \sum_{i=1}^n x_i y_i \right) - \bar{x} \bar{y}.$$

Coefficiente di correlazione o Indice di Pearson

Definizione (Coefficiente di correlazione o indice di Pearson)

Si dice coefficiente di correlazione delle due variabili x e y la quantità

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}.$$

Coefficiente di correlazione o Indice di Pearson

Definizione (Coefficiente di correlazione o indice di Pearson)

Si dice coefficiente di correlazione delle due variabili x e y la quantità

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}.$$

Teorema

Sia $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ un insieme di n osservazioni congiunte di due variabili. Allora

$$-1 \leq \rho_{xy} \leq 1.$$

Coefficiente di correlazione o Indice di Pearson

Teorema

$$-1 \leq \rho_{xy} \leq 1.$$

Dimostrazione

Coefficiente di correlazione o Indice di Pearson

Teorema

$$-1 \leq \rho_{xy} \leq 1.$$

Dimostrazione

Nota che

$$-1 \leq \rho_{xy} \leq 1 \Leftrightarrow |\rho_{xy}| \leq 1 \Leftrightarrow \frac{|\sigma_{xy}|}{\sigma_x \sigma_y} \leq 1 \Leftrightarrow |\sigma_{xy}| \leq \sigma_x \sigma_y.$$

Sostituendo le espressioni per la covarianza e la deviazione standard otteniamo:

$$\begin{aligned} |\sigma_{xy}| \leq \sigma_x \sigma_y &\Leftrightarrow \left| \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right| \leq \sqrt{\left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right] \left[\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \right]} \\ &\Leftrightarrow \left[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right]^2 \leq \left[\sum_{i=1}^n (x_i - \bar{x})^2 \right] \left[\sum_{i=1}^n (y_i - \bar{y})^2 \right]. \end{aligned}$$

Coefficiente di correlazione o Indice di Pearson

Teorema

$$-1 \leq \rho_{xy} \leq 1.$$

Dimostrazione

Quest'ultima disequazione può essere dimostrata per induzione su n .

Semplifichiamo l'espressione ponendo $\alpha_i = x_i - \bar{x}$ e $\beta_i = y_i - \bar{y}$.

Per $n = 1$ abbiamo

$$\left[\sum_{i=1}^1 \alpha_i \beta_i \right]^2 = (\alpha_1 \beta_1)^2 = \alpha_1^2 \beta_1^2 = \left[\sum_{i=1}^1 \alpha_i^2 \right] \left[\sum_{i=1}^1 \beta_i^2 \right].$$

Supponiamo che l'espressione sia vera per $n - 1$ e dimostriamo che vera per n .

$$\left[\sum_{i=1}^n \alpha_i \beta_i \right]^2 \leq \left[\sum_{i=1}^n \alpha_i^2 \right] \left[\sum_{i=1}^n \beta_i^2 \right] \Leftrightarrow$$

$$[(\alpha_1 \beta_1 + \dots + \alpha_{n-1} \beta_{n-1}) + \alpha_n \beta_n]^2 \leq [(\alpha_1^2 + \dots + \alpha_{n-1}^2) + \alpha_n^2] [(\beta_1^2 + \dots + \beta_{n-1}^2) + \beta_n^2].$$

Coefficiente di correlazione o Indice di Pearson

Teorema

$$-1 \leq \rho_{xy} \leq 1.$$

Dimostrazione

Applicando l'ipotesi induttiva, la precedente disequazione è equivalente a

$$2(\alpha_1\beta_1 + \dots + \alpha_{n-1}\beta_{n-1})\alpha_n\beta_n \leq (\alpha_1^2 + \dots + \alpha_{n-1}^2)\beta_n^2 + (\beta_1^2 + \dots + \beta_{n-1}^2)\alpha_n^2 \Leftrightarrow$$
$$(\alpha_1^2 + \dots + \alpha_{n-1}^2)\beta_n^2 + (\beta_1^2 + \dots + \beta_{n-1}^2)\alpha_n^2 - 2(\alpha_1\beta_1 + \dots + \alpha_{n-1}\beta_{n-1})\alpha_n\beta_n \geq 0.$$

A questo punto, attraverso opportune messe in evidenza, si ottiene che la precedente disequazione è equivalente a

$$(\alpha_1\beta_n - \beta_1\alpha_n)^2 + \dots + (\alpha_{n-1}\beta_n - \beta_{n-1}\alpha_n)^2 \geq 0.$$

Coefficiente di correlazione o Indice di Pearson

Le variabili x e y sono

- *direttamente correlate* se $\rho_{xy} \simeq 1$;
- *inversamente correlate* se $\rho_{xy} \simeq -1$;
- *incorrelate* se $\rho_{xy} \simeq 0$.

In modo più dettagliato possiamo dire che le variabili x e y sono

- *fortemente direttamente correlate* se $0.7 < \rho_{xy} \leq 1$;
- *moderatamente direttamente correlate* se $0.3 < \rho_{xy} \leq 0.7$;
- *scarsamente direttamente correlate* se $0 < \rho_{xy} \leq 0.3$;
- *scarsamente inversamente correlate* se $-0.3 \leq \rho_{xy} < 0$;
- *moderatamente inversamente correlate* se $-0.7 \leq \rho_{xy} < -0.3$;
- *fortemente inversamente correlate* se $-1 \leq \rho_{xy} < -0.7$;

Coefficiente di correlazione o Indice di Pearson

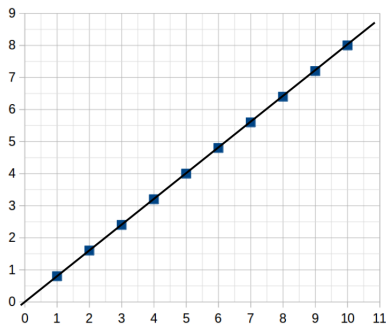
Osservazione

La variabile statistica Y è dipendente linearmente da X secondo la relazione $Y = a + bX$ se e solo se il coefficiente di correlazione vale 1 oppure -1 . In particolare vale 1 se $b > 0$ e vale -1 se $b < 0$.

Coefficiente di correlazione o Indice di Pearson

Osservazione

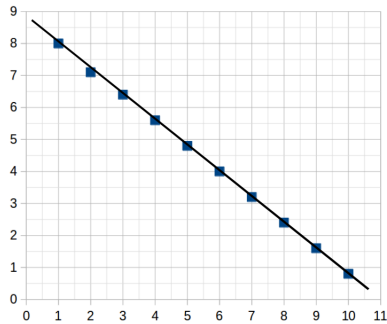
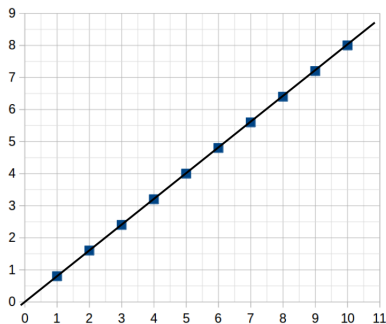
La variabile statistica Y è dipendente linearmente da X secondo la relazione $Y = a + bX$ se e solo se il coefficiente di correlazione vale 1 oppure -1 . In particolare vale 1 se $b > 0$ e vale -1 se $b < 0$.



Coefficiente di correlazione o Indice di Pearson

Osservazione

La variabile statistica Y è dipendente linearmente da X secondo la relazione $Y = a + bX$ se e solo se il coefficiente di correlazione vale 1 oppure -1 . In particolare vale 1 se $b > 0$ e vale -1 se $b < 0$.



Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707			
2	13.80	1137			
3	18.70	1451			
4	6.10	848			
5	6.20	315			
6	8.10	1130			
7	5.50	412			
8	10.20	732			
9	10.10	1893			
10	6.00	538			

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707			
2	13.80	1137			
3	18.70	1451			
4	6.10	848			
5	6.20	315			
6	8.10	1130			
7	5.50	412			
8	10.20	732			
9	10.10	1893			
10	6.00	538			
$\sum_{i=1}^{10}$					

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707			
2	13.80	1137			
3	18.70	1451			
4	6.10	848			
5	6.20	315			
6	8.10	1130			
7	5.50	412			
8	10.20	732			
9	10.10	1893			
10	6.00	538			
$\sum_{i=1}^{10}$	108.3				

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707			
2	13.80	1137			
3	18.70	1451			
4	6.10	848			
5	6.20	315			
6	8.10	1130			
7	5.50	412			
8	10.20	732			
9	10.10	1893			
10	6.00	538			
$\sum_{i=1}^{10}$	108.3	11163			

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2		
2	13.80	1137			
3	18.70	1451			
4	6.10	848			
5	6.20	315			
6	8.10	1130			
7	5.50	412			
8	10.20	732			
9	10.10	1893			
10	6.00	538			
$\sum_{i=1}^{10}$	108.3	11163			

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2		
2	13.80	1137	15690.6		
3	18.70	1451			
4	6.10	848			
5	6.20	315			
6	8.10	1130			
7	5.50	412			
8	10.20	732			
9	10.10	1893			
10	6.00	538			
$\sum_{i=1}^{10}$	108.3	11163			

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2		
2	13.80	1137	15690.6		
3	18.70	1451	27133.7		
4	6.10	848	5172.8		
5	6.20	315	1953.0		
6	8.10	1130	9153.0		
7	5.50	412	2266.0		
8	10.20	732	7466.4		
9	10.10	1893	19119.3		
10	6.00	538	3228.0		
$\sum_{i=1}^{10}$	108.3	11163			

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	
2	13.80	1137	15690.6		
3	18.70	1451	27133.7		
4	6.10	848	5172.8		
5	6.20	315	1953.0		
6	8.10	1130	9153.0		
7	5.50	412	2266.0		
8	10.20	732	7466.4		
9	10.10	1893	19119.3		
10	6.00	538	3228.0		
$\sum_{i=1}^{10}$	108.3	11163			

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	
2	13.80	1137	15690.6	190.44	
3	18.70	1451	27133.7		
4	6.10	848	5172.8		
5	6.20	315	1953.0		
6	8.10	1130	9153.0		
7	5.50	412	2266.0		
8	10.20	732	7466.4		
9	10.10	1893	19119.3		
10	6.00	538	3228.0		
$\sum_{i=1}^{10}$	108.3	11163			

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	
2	13.80	1137	15690.6	190.44	
3	18.70	1451	27133.7	349.69	
4	6.10	848	5172.8	37.21	
5	6.20	315	1953.0	38.44	
6	8.10	1130	9153.0	65.61	
7	5.50	412	2266.0	30.25	
8	10.20	732	7466.4	104.04	
9	10.10	1893	19119.3	102.01	
10	6.00	538	3228.0	36.00	
$\sum_{i=1}^{10}$	108.3	11163			

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	
3	18.70	1451	27133.7	349.69	
4	6.10	848	5172.8	37.21	
5	6.20	315	1953.0	38.44	
6	8.10	1130	9153.0	65.61	
7	5.50	412	2266.0	30.25	
8	10.20	732	7466.4	104.04	
9	10.10	1893	19119.3	102.01	
10	6.00	538	3228.0	36.00	
$\sum_{i=1}^{10}$	108.3	11163			

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	
3	18.70	1451	27133.7	349.69	
4	6.10	848	5172.8	37.21	
5	6.20	315	1953.0	38.44	
6	8.10	1130	9153.0	65.61	
7	5.50	412	2266.0	30.25	
8	10.20	732	7466.4	104.04	
9	10.10	1893	19119.3	102.01	
10	6.00	538	3228.0	36.00	
$\sum_{i=1}^{10}$	108.3	11163			

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	1292769
3	18.70	1451	27133.7	349.69	2105401
4	6.10	848	5172.8	37.21	719104
5	6.20	315	1953.0	38.44	99225
6	8.10	1130	9153.0	65.61	1276900
7	5.50	412	2266.0	30.25	169744
8	10.20	732	7466.4	104.04	535824
9	10.10	1893	19119.3	102.01	3583449
10	6.00	538	3228.0	36.00	289444
$\sum_{i=1}^{10}$	108.3	11163			

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	1292769
3	18.70	1451	27133.7	349.69	2105401
4	6.10	848	5172.8	37.21	719104
5	6.20	315	1953.0	38.44	99225
6	8.10	1130	9153.0	65.61	1276900
7	5.50	412	2266.0	30.25	169744
8	10.20	732	7466.4	104.04	535824
9	10.10	1893	19119.3	102.01	3583449
10	6.00	538	3228.0	36.00	289444
$\sum_{i=1}^{10}$	108.3	11163	155068		

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	1292769
3	18.70	1451	27133.7	349.69	2105401
4	6.10	848	5172.8	37.21	719104
5	6.20	315	1953.0	38.44	99225
6	8.10	1130	9153.0	65.61	1276900
7	5.50	412	2266.0	30.25	169744
8	10.20	732	7466.4	104.04	535824
9	10.10	1893	19119.3	102.01	3583449
10	6.00	538	3228.0	36.00	289444
$\sum_{i=1}^{10}$	108.3	11163	155068	1510.65	

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	1292769
3	18.70	1451	27133.7	349.69	2105401
4	6.10	848	5172.8	37.21	719104
5	6.20	315	1953.0	38.44	99225
6	8.10	1130	9153.0	65.61	1276900
7	5.50	412	2266.0	30.25	169744
8	10.20	732	7466.4	104.04	535824
9	10.10	1893	19119.3	102.01	3583449
10	6.00	538	3228.0	36.00	289444
$\sum_{i=1}^{10}$	108.3	11163	155068	1510.65	17399709

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	1292769
3	18.70	1451	27133.7	349.69	2105401
4	6.10	848	5172.8	37.21	719104
5	6.20	315	1953.0	38.44	99225
6	8.10	1130	9153.0	65.61	1276900
7	5.50	412	2266.0	30.25	169744
8	10.20	732	7466.4	104.04	535824
9	10.10	1893	19119.3	102.01	3583449
10	6.00	538	3228.0	36.00	289444
$\sum_{i=1}^{10}$	108.3	11163	155068	1510.65	17399709

Calcoliamo le medie: $\bar{x} = \frac{108.3}{10} = 10.83$; $\bar{y} = \frac{11163}{10} = 1116.3$.

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	1292769
3	18.70	1451	27133.7	349.69	2105401
4	6.10	848	5172.8	37.21	719104
5	6.20	315	1953.0	38.44	99225
6	8.10	1130	9153.0	65.61	1276900
7	5.50	412	2266.0	30.25	169744
8	10.20	732	7466.4	104.04	535824
9	10.10	1893	19119.3	102.01	3583449
10	6.00	538	3228.0	36.00	289444
$\sum_{i=1}^{10}$	108.3	11163	155068	1510.65	17399709

Calcoliamo le **medie**: $\bar{x} = \frac{108.3}{10} = 10.83$; $\bar{y} = \frac{11163}{10} = 1116.3$.

Calcoliamo le **varianze**: $\sigma_x^2 = \frac{1510.65}{10} - 10.83^2$

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	1292769
3	18.70	1451	27133.7	349.69	2105401
4	6.10	848	5172.8	37.21	719104
5	6.20	315	1953.0	38.44	99225
6	8.10	1130	9153.0	65.61	1276900
7	5.50	412	2266.0	30.25	169744
8	10.20	732	7466.4	104.04	535824
9	10.10	1893	19119.3	102.01	3583449
10	6.00	538	3228.0	36.00	289444
$\sum_{i=1}^{10}$	108.3	11163	155068	1510.65	17399709

Calcoliamo le **medie**: $\bar{x} = \frac{108.3}{10} = 10.83$; $\bar{y} = \frac{11163}{10} = 1116.3$.

Calcoliamo le **varianze**: $\sigma_x^2 = \frac{1510.65}{10} - 10.83^2 = 33.7761$;

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	1292769
3	18.70	1451	27133.7	349.69	2105401
4	6.10	848	5172.8	37.21	719104
5	6.20	315	1953.0	38.44	99225
6	8.10	1130	9153.0	65.61	1276900
7	5.50	412	2266.0	30.25	169744
8	10.20	732	7466.4	104.04	535824
9	10.10	1893	19119.3	102.01	3583449
10	6.00	538	3228.0	36.00	289444
$\sum_{i=1}^{10}$	108.3	11163	155068	1510.65	17399709

Calcoliamo le **medie**: $\bar{x} = \frac{108.3}{10} = 10.83$; $\bar{y} = \frac{11163}{10} = 1116.3$.

Calcoliamo le **varianze**: $\sigma_x^2 = \frac{1510.65}{10} - 10.83^2 = 33.7761$; $\sigma_y^2 = \frac{17399709}{10} - 1116.3^2$

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	1292769
3	18.70	1451	27133.7	349.69	2105401
4	6.10	848	5172.8	37.21	719104
5	6.20	315	1953.0	38.44	99225
6	8.10	1130	9153.0	65.61	1276900
7	5.50	412	2266.0	30.25	169744
8	10.20	732	7466.4	104.04	535824
9	10.10	1893	19119.3	102.01	3583449
10	6.00	538	3228.0	36.00	289444
$\sum_{i=1}^{10}$	108.3	11163	155068	1510.65	17399709

Calcoliamo le **medie**: $\bar{x} = \frac{108.3}{10} = 10.83$; $\bar{y} = \frac{11163}{10} = 1116.3$.

Calcoliamo le **varianze**: $\sigma_x^2 = \frac{1510.65}{10} - 10.83^2 = 33.7761$; $\sigma_y^2 = \frac{17399709}{10} - 1116.3^2 = 493845.2$.

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	1292769
3	18.70	1451	27133.7	349.69	2105401
4	6.10	848	5172.8	37.21	719104
5	6.20	315	1953.0	38.44	99225
6	8.10	1130	9153.0	65.61	1276900
7	5.50	412	2266.0	30.25	169744
8	10.20	732	7466.4	104.04	535824
9	10.10	1893	19119.3	102.01	3583449
10	6.00	538	3228.0	36.00	289444
$\sum_{i=1}^{10}$	108.3	11163	155068	1510.65	17399709

Calcoliamo le **medie**: $\bar{x} = \frac{108.3}{10} = 10.83$; $\bar{y} = \frac{11163}{10} = 1116.3$.

Calcoliamo le **varianze**: $\sigma_x^2 = \frac{1510.65}{10} - 10.83^2 = 33.7761$; $\sigma_y^2 = \frac{17399709}{10} - 1116.3^2 = 493845.2$.

Calcoliamo le **deviazioni standard**:

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	1292769
3	18.70	1451	27133.7	349.69	2105401
4	6.10	848	5172.8	37.21	719104
5	6.20	315	1953.0	38.44	99225
6	8.10	1130	9153.0	65.61	1276900
7	5.50	412	2266.0	30.25	169744
8	10.20	732	7466.4	104.04	535824
9	10.10	1893	19119.3	102.01	3583449
10	6.00	538	3228.0	36.00	289444
$\sum_{i=1}^{10}$	108.3	11163	155068	1510.65	17399709

Calcoliamo le **medie**: $\bar{x} = \frac{108.3}{10} = 10.83$; $\bar{y} = \frac{11163}{10} = 1116.3$.

Calcoliamo le **varianze**: $\sigma_x^2 = \frac{1510.65}{10} - 10.83^2 = 33.7761$; $\sigma_y^2 = \frac{17399709}{10} - 1116.3^2 = 493845.2$.

Calcoliamo le **deviazioni standard**:

$$\sigma_x = \sqrt{\sigma_x^2}$$

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	1292769
3	18.70	1451	27133.7	349.69	2105401
4	6.10	848	5172.8	37.21	719104
5	6.20	315	1953.0	38.44	99225
6	8.10	1130	9153.0	65.61	1276900
7	5.50	412	2266.0	30.25	169744
8	10.20	732	7466.4	104.04	535824
9	10.10	1893	19119.3	102.01	3583449
10	6.00	538	3228.0	36.00	289444
$\sum_{i=1}^{10}$	108.3	11163	155068	1510.65	17399709

Calcoliamo le **medie**: $\bar{x} = \frac{108.3}{10} = 10.83$; $\bar{y} = \frac{11163}{10} = 1116.3$.

Calcoliamo le **varianze**: $\sigma_x^2 = \frac{1510.65}{10} - 10.83^2 = 33.7761$; $\sigma_y^2 = \frac{17399709}{10} - 1116.3^2 = 493845.2$.

Calcoliamo le **deviazioni standard**:

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{33.7761} \simeq 5.811721;$$

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	1292769
3	18.70	1451	27133.7	349.69	2105401
4	6.10	848	5172.8	37.21	719104
5	6.20	315	1953.0	38.44	99225
6	8.10	1130	9153.0	65.61	1276900
7	5.50	412	2266.0	30.25	169744
8	10.20	732	7466.4	104.04	535824
9	10.10	1893	19119.3	102.01	3583449
10	6.00	538	3228.0	36.00	289444
$\sum_{i=1}^{10}$	108.3	11163	155068	1510.65	17399709

Calcoliamo le **medie**: $\bar{x} = \frac{108.3}{10} = 10.83$; $\bar{y} = \frac{11163}{10} = 1116.3$.

Calcoliamo le **varianze**: $\sigma_x^2 = \frac{1510.65}{10} - 10.83^2 = 33.7761$; $\sigma_y^2 = \frac{17399709}{10} - 1116.3^2 = 493845.2$.

Calcoliamo le **deviazioni standard**:

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{33.7761} \simeq 5.811721; \sigma_y = \sqrt{\sigma_y^2}$$

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	1292769
3	18.70	1451	27133.7	349.69	2105401
4	6.10	848	5172.8	37.21	719104
5	6.20	315	1953.0	38.44	99225
6	8.10	1130	9153.0	65.61	1276900
7	5.50	412	2266.0	30.25	169744
8	10.20	732	7466.4	104.04	535824
9	10.10	1893	19119.3	102.01	3583449
10	6.00	538	3228.0	36.00	289444
$\sum_{i=1}^{10}$	108.3	11163	155068	1510.65	17399709

Calcoliamo le **medie**: $\bar{x} = \frac{108.3}{10} = 10.83$; $\bar{y} = \frac{11163}{10} = 1116.3$.

Calcoliamo le **varianze**: $\sigma_x^2 = \frac{1510.65}{10} - 10.83^2 = 33.7761$; $\sigma_y^2 = \frac{17399709}{10} - 1116.3^2 = 493845.2$.

Calcoliamo le **deviazioni standard**:

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{33.7761} \simeq 5.811721; \sigma_y = \sqrt{\sigma_y^2} = \sqrt{493845.2} \simeq 702.7412.$$

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	1292769
3	18.70	1451	27133.7	349.69	2105401
4	6.10	848	5172.8	37.21	719104
5	6.20	315	1953.0	38.44	99225
6	8.10	1130	9153.0	65.61	1276900
7	5.50	412	2266.0	30.25	169744
8	10.20	732	7466.4	104.04	535824
9	10.10	1893	19119.3	102.01	3583449
10	6.00	538	3228.0	36.00	289444
$\sum_{i=1}^{10}$	108.3	11163	155068	1510.65	17399709

Calcoliamo le **medie**: $\bar{x} = \frac{108.3}{10} = 10.83$; $\bar{y} = \frac{11163}{10} = 1116.3$.

Calcoliamo le **varianze**: $\sigma_x^2 = \frac{1510.65}{10} - 10.83^2 = 33.7761$; $\sigma_y^2 = \frac{17399709}{10} - 1116.3^2 = 493845.2$.

Calcoliamo le **deviazioni standard**:

$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{33.7761} \simeq 5.811721$; $\sigma_y = \sqrt{\sigma_y^2} = \sqrt{493845.2} \simeq 702.7412$.

Calcoliamo la **covarianza**:

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	1292769
3	18.70	1451	27133.7	349.69	2105401
4	6.10	848	5172.8	37.21	719104
5	6.20	315	1953.0	38.44	99225
6	8.10	1130	9153.0	65.61	1276900
7	5.50	412	2266.0	30.25	169744
8	10.20	732	7466.4	104.04	535824
9	10.10	1893	19119.3	102.01	3583449
10	6.00	538	3228.0	36.00	289444
$\sum_{i=1}^{10}$	108.3	11163	155068	1510.65	17399709

Calcoliamo le **medie**: $\bar{x} = \frac{108.3}{10} = 10.83$; $\bar{y} = \frac{11163}{10} = 1116.3$.

Calcoliamo le **varianze**: $\sigma_x^2 = \frac{1510.65}{10} - 10.83^2 = 33.7761$; $\sigma_y^2 = \frac{17399709}{10} - 1116.3^2 = 493845.2$.

Calcoliamo le **deviazioni standard**:

$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{33.7761} \simeq 5.811721$; $\sigma_y = \sqrt{\sigma_y^2} = \sqrt{493845.2} \simeq 702.7412$.

Calcoliamo la **covarianza**:

$$\sigma_{xy} = \frac{155068}{10} - 10.83 \cdot 1116.3$$

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	1292769
3	18.70	1451	27133.7	349.69	2105401
4	6.10	848	5172.8	37.21	719104
5	6.20	315	1953.0	38.44	99225
6	8.10	1130	9153.0	65.61	1276900
7	5.50	412	2266.0	30.25	169744
8	10.20	732	7466.4	104.04	535824
9	10.10	1893	19119.3	102.01	3583449
10	6.00	538	3228.0	36.00	289444
$\sum_{i=1}^{10}$	108.3	11163	155068	1510.65	17399709

Calcoliamo le **medie**: $\bar{x} = \frac{108.3}{10} = 10.83$; $\bar{y} = \frac{11163}{10} = 1116.3$.

Calcoliamo le **varianze**: $\sigma_x^2 = \frac{1510.65}{10} - 10.83^2 = 33.7761$; $\sigma_y^2 = \frac{17399709}{10} - 1116.3^2 = 493845.2$.

Calcoliamo le **deviazioni standard**:

$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{33.7761} \simeq 5.811721$; $\sigma_y = \sqrt{\sigma_y^2} = \sqrt{493845.2} \simeq 702.7412$.

Calcoliamo la **covarianza**:

$$\sigma_{xy} = \frac{155068}{10} - 10.83 \cdot 1116.3 = 3417.271.$$

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	1292769
3	18.70	1451	27133.7	349.69	2105401
4	6.10	848	5172.8	37.21	719104
5	6.20	315	1953.0	38.44	99225
6	8.10	1130	9153.0	65.61	1276900
7	5.50	412	2266.0	30.25	169744
8	10.20	732	7466.4	104.04	535824
9	10.10	1893	19119.3	102.01	3583449
10	6.00	538	3228.0	36.00	289444
$\sum_{i=1}^{10}$	108.3	11163	155068	1510.65	17399709

Calcoliamo le **medie**: $\bar{x} = \frac{108.3}{10} = 10.83$; $\bar{y} = \frac{11163}{10} = 1116.3$.

Calcoliamo le **varianze**: $\sigma_x^2 = \frac{1510.65}{10} - 10.83^2 = 33.7761$; $\sigma_y^2 = \frac{17399709}{10} - 1116.3^2 = 493845.2$.

Calcoliamo le **deviazioni standard**:

$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{33.7761} \simeq 5.811721$; $\sigma_y = \sqrt{\sigma_y^2} = \sqrt{493845.2} \simeq 702.7412$.

Calcoliamo la **covarianza**:

$$\sigma_{xy} = \frac{155068}{10} - 10.83 \cdot 1116.3 = 3417.271.$$

Il coefficiente di correlazione è: $\rho = \frac{3417.271}{5.811721 \cdot 702.7412}$

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	1292769
3	18.70	1451	27133.7	349.69	2105401
4	6.10	848	5172.8	37.21	719104
5	6.20	315	1953.0	38.44	99225
6	8.10	1130	9153.0	65.61	1276900
7	5.50	412	2266.0	30.25	169744
8	10.20	732	7466.4	104.04	535824
9	10.10	1893	19119.3	102.01	3583449
10	6.00	538	3228.0	36.00	289444
$\sum_{i=1}^{10}$	108.3	11163	155068	1510.65	17399709

Calcoliamo le **medie**: $\bar{x} = \frac{108.3}{10} = 10.83$; $\bar{y} = \frac{11163}{10} = 1116.3$.

Calcoliamo le **varianze**: $\sigma_x^2 = \frac{1510.65}{10} - 10.83^2 = 33.7761$; $\sigma_y^2 = \frac{17399709}{10} - 1116.3^2 = 493845.2$.

Calcoliamo le **deviazioni standard**:

$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{33.7761} \simeq 5.811721$; $\sigma_y = \sqrt{\sigma_y^2} = \sqrt{493845.2} \simeq 702.7412$.

Calcoliamo la **covarianza**:

$$\sigma_{xy} = \frac{155068}{10} - 10.83 \cdot 1116.3 = 3417.271.$$

Il coefficiente di correlazione è: $\rho = \frac{3417.271}{5.811721 \cdot 702.7412} \simeq 0.84$.

Coefficiente di correlazione o Indice di Pearson

Esempio (Consideriamo i dati dell'Esempio 23)

Costruiamo la seguente tabella che sarà utile per calcolare i vari indici.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	23.60	2707	63885.2	556.96	7327849
2	13.80	1137	15690.6	190.44	1292769
3	18.70	1451	27133.7	349.69	2105401
4	6.10	848	5172.8	37.21	719104
5	6.20	315	1953.0	38.44	99225
6	8.10	1130	9153.0	65.61	1276900
7	5.50	412	2266.0	30.25	169744
8	10.20	732	7466.4	104.04	535824
9	10.10	1893	19119.3	102.01	3583449
10	6.00	538	3228.0	36.00	289444
$\sum_{i=1}^{10}$	108.3	11163	155068	1510.65	17399709

Calcoliamo le **medie**: $\bar{x} = \frac{108.3}{10} = 10.83$; $\bar{y} = \frac{11163}{10} = 1116.3$.

Calcoliamo le **varianze**: $\sigma_x^2 = \frac{1510.65}{10} - 10.83^2 = 33.7761$; $\sigma_y^2 = \frac{17399709}{10} - 1116.3^2 = 493845.2$.

Calcoliamo le **deviazioni standard**:

$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{33.7761} \simeq 5.811721$; $\sigma_y = \sqrt{\sigma_y^2} = \sqrt{493845.2} \simeq 702.7412$.

Calcoliamo la **covarianza**:

$$\sigma_{xy} = \frac{155068}{10} - 10.83 \cdot 1116.3 = 3417.271.$$

Il coefficiente di correlazione è: $\rho = \frac{3417.271}{5.811721 \cdot 702.7412} \simeq 0.84$. Tra i dati esiste una **forte correlazione diretta**.

Retta di regressione (Metodo dei minimi quadrati)

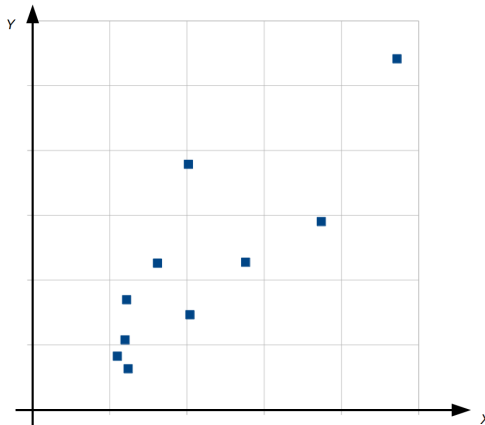
Fine Trovare una retta che approssima meglio i punti in uno scatterplot.

Mezzo Il metodo dei minimi quadrati consiste nel determinare la retta che minimizza la somma dei quadrati delle distanze tra essa e i punti.

Retta di regressione (Metodo dei minimi quadrati)

Fine Trovare una retta che approssima meglio i punti in uno scatterplot.

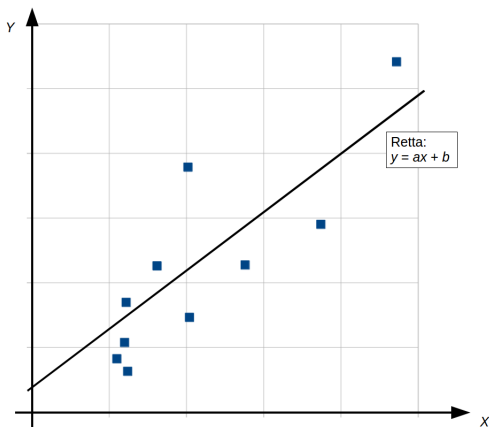
Mezzo Il metodo dei minimi quadrati consiste nel determinare la retta che minimizza la somma dei quadrati delle distanze tra essa e i punti.



Retta di regressione (Metodo dei minimi quadrati)

Fine Trovare una retta che approssima meglio i punti in uno scatterplot.

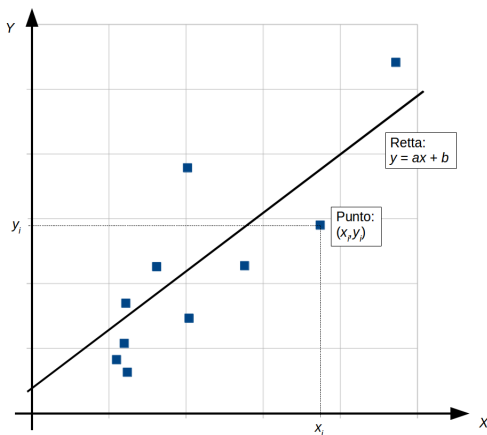
Mezzo Il metodo dei minimi quadrati consiste nel determinare la retta che minimizza la somma dei quadrati delle distanze tra essa e i punti.



Retta di regressione (Metodo dei minimi quadrati)

Fine Trovare una retta che approssima meglio i punti in uno scatterplot.

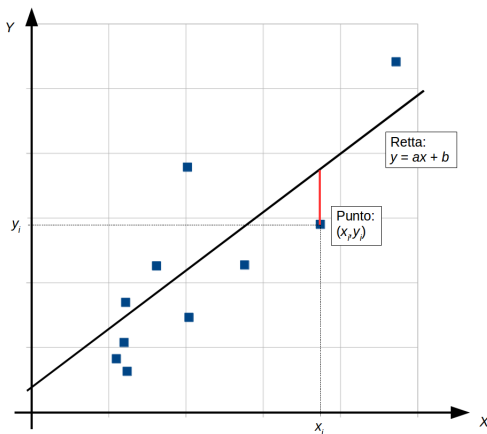
Mezzo Il metodo dei minimi quadrati consiste nel determinare la retta che minimizza la somma dei quadrati delle distanze tra essa e i punti.



Retta di regressione (Metodo dei minimi quadrati)

Fine Trovare una retta che approssima meglio i punti in uno scatterplot.

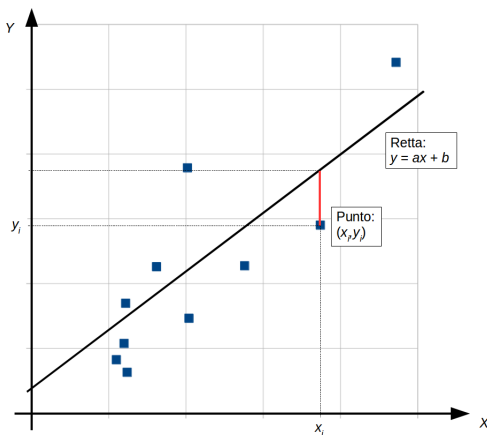
Mezzo Il metodo dei minimi quadrati consiste nel determinare la retta che minimizza la somma dei quadrati delle distanze tra essa e i punti.



Retta di regressione (Metodo dei minimi quadrati)

Fine Trovare una retta che approssima meglio i punti in uno scatterplot.

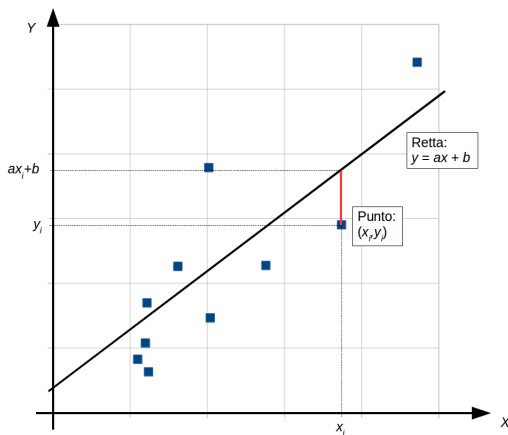
Mezzo Il metodo dei minimi quadrati consiste nel determinare la retta che minimizza la somma dei quadrati delle distanze tra essa e i punti.



Retta di regressione (Metodo dei minimi quadrati)

Fine Trovare una retta che approssima meglio i punti in uno scatterplot.

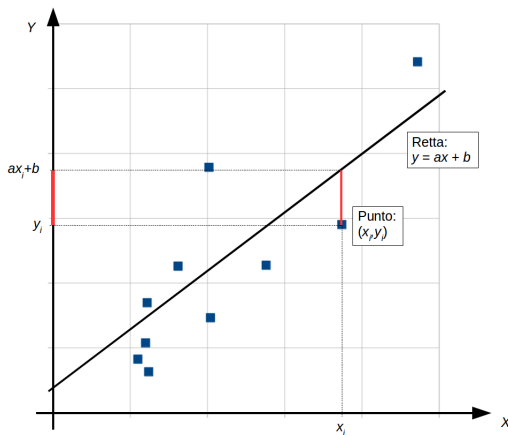
Mezzo Il metodo dei minimi quadrati consiste nel determinare la retta che minimizza la somma dei quadrati delle distanze tra essa e i punti.



Retta di regressione (Metodo dei minimi quadrati)

Fine Trovare una retta che approssima meglio i punti in uno scatterplot.

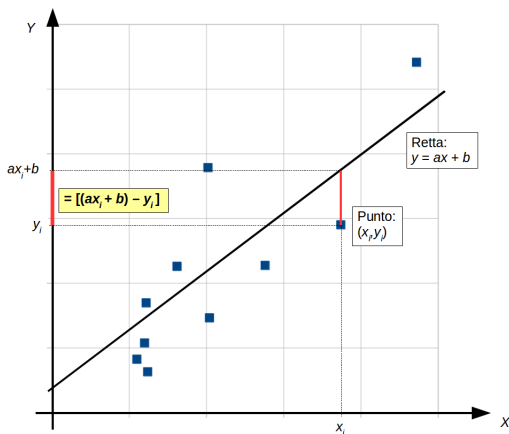
Mezzo Il metodo dei minimi quadrati consiste nel determinare la retta che minimizza la somma dei quadrati delle distanze tra essa e i punti.



Retta di regressione (Metodo dei minimi quadrati)

Fine Trovare una retta che approssima meglio i punti in uno scatterplot.

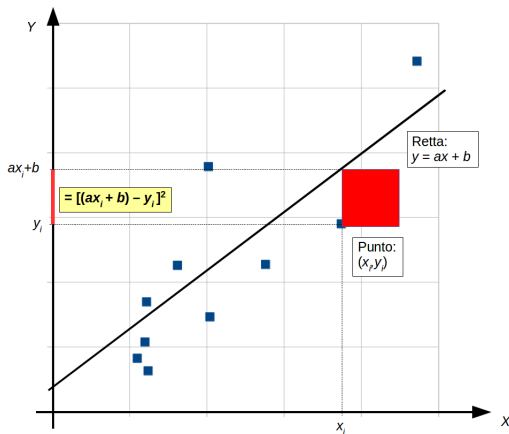
Mezzo Il metodo dei minimi quadrati consiste nel determinare la retta che minimizza la somma dei quadrati delle distanze tra essa e i punti.



Retta di regressione (Metodo dei minimi quadrati)

Fine Trovare una retta che approssima meglio i punti in uno scatterplot.

Mezzo Il metodo dei minimi quadrati consiste nel determinare la retta che minimizza la somma dei quadrati delle distanze tra essa e i punti.



Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione.

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

{

}

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

$$\left\{ \begin{array}{l} \frac{\partial f}{\partial a}(a, b) \\ \frac{\partial f}{\partial b}(a, b) \end{array} \right.$$

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

$$\begin{cases} \frac{\partial f}{\partial a}(a, b) = - \sum_{i=1}^n 2x_i [y_i - (ax_i + b)] \\ \frac{\partial f}{\partial b}(a, b) = - \sum_{i=1}^n 2 [y_i - (ax_i + b)] \end{cases}$$

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

$$\left\{ \begin{array}{l} \frac{\partial f}{\partial a}(a, b) = - \sum_{i=1}^n 2x_i [y_i - (ax_i + b)] = 0 \end{array} \right.$$

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

$$\begin{cases} \frac{\partial f}{\partial a}(a, b) = - \sum_{i=1}^n 2x_i [y_i - (ax_i + b)] = 0 \\ \frac{\partial f}{\partial b}(a, b) \end{cases}$$

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

$$\begin{cases} \frac{\partial f}{\partial a}(a, b) = - \sum_{i=1}^n 2x_i [y_i - (ax_i + b)] = 0 \\ \frac{\partial f}{\partial b}(a, b) = - \sum_{i=1}^n 2 [y_i - (ax_i + b)] \end{cases}$$

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

$$\begin{cases} \frac{\partial f}{\partial a}(a, b) = - \sum_{i=1}^n 2x_i [y_i - (ax_i + b)] = 0 \\ \frac{\partial f}{\partial b}(a, b) = - \sum_{i=1}^n 2 [y_i - (ax_i + b)] = 0 \end{cases}$$

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

$$\begin{cases} \frac{\partial f}{\partial a}(a, b) = - \sum_{i=1}^n 2x_i [y_i - (ax_i + b)] = 0 \\ \frac{\partial f}{\partial b}(a, b) = - \sum_{i=1}^n 2 [y_i - (ax_i + b)] = 0 \end{cases}$$

Dalla seconda equazione otteniamo $b = \frac{1}{n} \sum_{i=1}^n (y_i - ax_i).$

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

$$\begin{cases} \frac{\partial f}{\partial a}(a, b) = -\sum_{i=1}^n 2x_i [y_i - (ax_i + b)] = 0 \\ \frac{\partial f}{\partial b}(a, b) = -\sum_{i=1}^n 2 [y_i - (ax_i + b)] = 0 \end{cases}$$

Dalla seconda equazione otteniamo $b = \frac{1}{n} \sum_{i=1}^n (y_i - ax_i)$. Sostituendo nella prima, abbiamo

$$\sum_{i=1}^n 2x_i \left[y_i - \left(ax_i + \frac{1}{n} \sum_{j=1}^n (y_j - ax_j) \right) \right]$$

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

$$\begin{cases} \frac{\partial f}{\partial a}(a, b) = -\sum_{i=1}^n 2x_i [y_i - (ax_i + b)] = 0 \\ \frac{\partial f}{\partial b}(a, b) = -\sum_{i=1}^n 2 [y_i - (ax_i + b)] = 0 \end{cases}$$

Dalla seconda equazione otteniamo $b = \frac{1}{n} \sum_{i=1}^n (y_i - ax_i)$. Sostituendo nella prima, abbiamo

$$\sum_{i=1}^n 2x_i \left[y_i - \left(ax_i + \frac{1}{n} \sum_{j=1}^n (y_j - ax_j) \right) \right] = 2 \left[\sum_{i=1}^n x_i \left(y_i - \frac{1}{n} \sum_{j=1}^n y_j \right) - a \sum_{i=1}^n x_i \left(x_i - \frac{1}{n} \sum_{j=1}^n x_j \right) \right] =$$

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

$$\begin{cases} \frac{\partial f}{\partial a}(a, b) = -\sum_{i=1}^n 2x_i [y_i - (ax_i + b)] = 0 \\ \frac{\partial f}{\partial b}(a, b) = -\sum_{i=1}^n 2 [y_i - (ax_i + b)] = 0 \end{cases}$$

Dalla seconda equazione otteniamo $b = \frac{1}{n} \sum_{i=1}^n (y_i - ax_i)$. Sostituendo nella prima, abbiamo

$$\begin{aligned} \sum_{i=1}^n 2x_i \left[y_i - \left(ax_i + \frac{1}{n} \sum_{j=1}^n (y_j - ax_j) \right) \right] &= 2 \left[\sum_{i=1}^n x_i \left(y_i - \frac{1}{n} \sum_{j=1}^n y_j \right) - a \sum_{i=1}^n x_i \left(x_i - \frac{1}{n} \sum_{j=1}^n x_j \right) \right] = \\ &= 2 \left[\sum_{i=1}^n x_i (y_i - \bar{y}) - a \sum_{i=1}^n x_i (x_i - \bar{x}) \right] \end{aligned}$$

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

$$\begin{cases} \frac{\partial f}{\partial a}(a, b) = -\sum_{i=1}^n 2x_i [y_i - (ax_i + b)] = 0 \\ \frac{\partial f}{\partial b}(a, b) = -\sum_{i=1}^n 2 [y_i - (ax_i + b)] = 0 \end{cases}$$

Dalla seconda equazione otteniamo $b = \frac{1}{n} \sum_{i=1}^n (y_i - ax_i)$. Sostituendo nella prima, abbiamo

$$\begin{aligned} \sum_{i=1}^n 2x_i \left[y_i - \left(ax_i + \frac{1}{n} \sum_{j=1}^n (y_j - ax_j) \right) \right] &= 2 \left[\sum_{i=1}^n x_i \left(y_i - \frac{1}{n} \sum_{j=1}^n y_j \right) - a \sum_{i=1}^n x_i \left(x_i - \frac{1}{n} \sum_{j=1}^n x_j \right) \right] = \\ &= 2 \left[\sum_{i=1}^n x_i (y_i - \bar{y}) - a \sum_{i=1}^n x_i (x_i - \bar{x}) \right] = 0 \end{aligned}$$

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

$$\begin{cases} \frac{\partial f}{\partial a}(a, b) = -\sum_{i=1}^n 2x_i [y_i - (ax_i + b)] = 0 \\ \frac{\partial f}{\partial b}(a, b) = -\sum_{i=1}^n 2 [y_i - (ax_i + b)] = 0 \end{cases}$$

Dalla seconda equazione otteniamo $b = \frac{1}{n} \sum_{i=1}^n (y_i - ax_i)$. Sostituendo nella prima, abbiamo

$$\begin{aligned} \sum_{i=1}^n 2x_i \left[y_i - \left(ax_i + \frac{1}{n} \sum_{j=1}^n (y_j - ax_j) \right) \right] &= 2 \left[\sum_{i=1}^n x_i \left(y_i - \frac{1}{n} \sum_{j=1}^n y_j \right) - a \sum_{i=1}^n x_i \left(x_i - \frac{1}{n} \sum_{j=1}^n x_j \right) \right] = \\ &= 2 \left[\sum_{i=1}^n x_i (y_i - \bar{y}) - a \sum_{i=1}^n x_i (x_i - \bar{x}) \right] = 0 \end{aligned}$$

Per cui, $a =$

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

$$\begin{cases} \frac{\partial f}{\partial a}(a, b) = - \sum_{i=1}^n 2x_i [y_i - (ax_i + b)] = 0 \\ \frac{\partial f}{\partial b}(a, b) = - \sum_{i=1}^n 2 [y_i - (ax_i + b)] = 0 \end{cases}$$

Dalla seconda equazione otteniamo $b = \frac{1}{n} \sum_{i=1}^n (y_i - ax_i)$. Sostituendo nella prima, abbiamo

$$\begin{aligned} \sum_{i=1}^n 2x_i \left[y_i - \left(ax_i + \frac{1}{n} \sum_{j=1}^n (y_j - ax_j) \right) \right] &= 2 \left[\sum_{i=1}^n x_i \left(y_i - \frac{1}{n} \sum_{j=1}^n y_j \right) - a \sum_{i=1}^n x_i \left(x_i - \frac{1}{n} \sum_{j=1}^n x_j \right) \right] = \\ &= 2 \left[\sum_{i=1}^n x_i (y_i - \bar{y}) - a \sum_{i=1}^n x_i (x_i - \bar{x}) \right] = 0 \end{aligned}$$

Per cui, $a = \frac{\sum_{i=1}^n x_i (y_i - \bar{y})}{\sum_{i=1}^n x_i (x_i - \bar{x})}$

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

$$\begin{cases} \frac{\partial f}{\partial a}(a, b) = -\sum_{i=1}^n 2x_i [y_i - (ax_i + b)] = 0 \\ \frac{\partial f}{\partial b}(a, b) = -\sum_{i=1}^n 2 [y_i - (ax_i + b)] = 0 \end{cases}$$

Dalla seconda equazione otteniamo $b = \frac{1}{n} \sum_{i=1}^n (y_i - ax_i)$. Sostituendo nella prima, abbiamo

$$\begin{aligned} \sum_{i=1}^n 2x_i \left[y_i - \left(ax_i + \frac{1}{n} \sum_{j=1}^n (y_j - ax_j) \right) \right] &= 2 \left[\sum_{i=1}^n x_i \left(y_i - \frac{1}{n} \sum_{j=1}^n y_j \right) - a \sum_{i=1}^n x_i \left(x_i - \frac{1}{n} \sum_{j=1}^n x_j \right) \right] = \\ &= 2 \left[\sum_{i=1}^n x_i (y_i - \bar{y}) - a \sum_{i=1}^n x_i (x_i - \bar{x}) \right] = 0 \end{aligned}$$

Per cui, $a = \frac{\sum_{i=1}^n x_i (y_i - \bar{y})}{\sum_{i=1}^n x_i (x_i - \bar{x})} = \frac{\sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i}$

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

$$\begin{cases} \frac{\partial f}{\partial a}(a, b) = -\sum_{i=1}^n 2x_i [y_i - (ax_i + b)] = 0 \\ \frac{\partial f}{\partial b}(a, b) = -\sum_{i=1}^n 2 [y_i - (ax_i + b)] = 0 \end{cases}$$

Dalla seconda equazione otteniamo $b = \frac{1}{n} \sum_{i=1}^n (y_i - ax_i)$. Sostituendo nella prima, abbiamo

$$\begin{aligned} \sum_{i=1}^n 2x_i \left[y_i - \left(ax_i + \frac{1}{n} \sum_{j=1}^n (y_j - ax_j) \right) \right] &= 2 \left[\sum_{i=1}^n x_i \left(y_i - \frac{1}{n} \sum_{j=1}^n y_j \right) - a \sum_{i=1}^n x_i \left(x_i - \frac{1}{n} \sum_{j=1}^n x_j \right) \right] = \\ &= 2 \left[\sum_{i=1}^n x_i (y_i - \bar{y}) - a \sum_{i=1}^n x_i (x_i - \bar{x}) \right] = 0 \end{aligned}$$

$$\text{Per cui, } a = \frac{\sum_{i=1}^n x_i (y_i - \bar{y})}{\sum_{i=1}^n x_i (x_i - \bar{x})} = \frac{\sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i} = \frac{n(\sigma_{xy} + \bar{x}\bar{y}) - \bar{y}(n\bar{x})}{n(\sigma_x^2 + \bar{x}^2) - \bar{x}(n\bar{x})}$$

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

$$\begin{cases} \frac{\partial f}{\partial a}(a, b) = -\sum_{i=1}^n 2x_i [y_i - (ax_i + b)] = 0 \\ \frac{\partial f}{\partial b}(a, b) = -\sum_{i=1}^n 2 [y_i - (ax_i + b)] = 0 \end{cases}$$

Dalla seconda equazione otteniamo $b = \frac{1}{n} \sum_{i=1}^n (y_i - ax_i)$. Sostituendo nella prima, abbiamo

$$\begin{aligned} \sum_{i=1}^n 2x_i \left[y_i - \left(ax_i + \frac{1}{n} \sum_{j=1}^n (y_j - ax_j) \right) \right] &= 2 \left[\sum_{i=1}^n x_i \left(y_i - \frac{1}{n} \sum_{j=1}^n y_j \right) - a \sum_{i=1}^n x_i \left(x_i - \frac{1}{n} \sum_{j=1}^n x_j \right) \right] = \\ &= 2 \left[\sum_{i=1}^n x_i (y_i - \bar{y}) - a \sum_{i=1}^n x_i (x_i - \bar{x}) \right] = 0 \end{aligned}$$

$$\text{Per cui, } a = \frac{\sum_{i=1}^n x_i (y_i - \bar{y})}{\sum_{i=1}^n x_i (x_i - \bar{x})} = \frac{\sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i} = \frac{n(\sigma_{xy} + \bar{x}\bar{y}) - \bar{y}(n\bar{x})}{n(\sigma_x^2 + \bar{x}^2) - \bar{x}(n\bar{x})} = \frac{\sigma_{xy}}{\sigma_x^2}$$

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

$$\begin{cases} \frac{\partial f}{\partial a}(a, b) = -\sum_{i=1}^n 2x_i [y_i - (ax_i + b)] = 0 \\ \frac{\partial f}{\partial b}(a, b) = -\sum_{i=1}^n 2 [y_i - (ax_i + b)] = 0 \end{cases}$$

Dalla seconda equazione otteniamo $b = \frac{1}{n} \sum_{i=1}^n (y_i - ax_i)$. Sostituendo nella prima, abbiamo

$$\begin{aligned} \sum_{i=1}^n 2x_i \left[y_i - \left(ax_i + \frac{1}{n} \sum_{j=1}^n (y_j - ax_j) \right) \right] &= 2 \left[\sum_{i=1}^n x_i \left(y_i - \frac{1}{n} \sum_{j=1}^n y_j \right) - a \sum_{i=1}^n x_i \left(x_i - \frac{1}{n} \sum_{j=1}^n x_j \right) \right] = \\ &= 2 \left[\sum_{i=1}^n x_i (y_i - \bar{y}) - a \sum_{i=1}^n x_i (x_i - \bar{x}) \right] = 0 \end{aligned}$$

$$\text{Per cui, } a = \frac{\sum_{i=1}^n x_i (y_i - \bar{y})}{\sum_{i=1}^n x_i (x_i - \bar{x})} = \frac{\sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i} = \frac{n(\sigma_{xy} + \bar{x}\bar{y}) - \bar{y}(n\bar{x})}{n(\sigma_x^2 + \bar{x}^2) - \bar{x}(n\bar{x})} = \frac{\sigma_{xy}}{\sigma_x^2} \text{ e}$$

$b =$

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

$$\begin{cases} \frac{\partial f}{\partial a}(a, b) = -\sum_{i=1}^n 2x_i [y_i - (ax_i + b)] = 0 \\ \frac{\partial f}{\partial b}(a, b) = -\sum_{i=1}^n 2 [y_i - (ax_i + b)] = 0 \end{cases}$$

Dalla seconda equazione otteniamo $b = \frac{1}{n} \sum_{i=1}^n (y_i - ax_i)$. Sostituendo nella prima, abbiamo

$$\begin{aligned} \sum_{i=1}^n 2x_i \left[y_i - \left(ax_i + \frac{1}{n} \sum_{j=1}^n (y_j - ax_j) \right) \right] &= 2 \left[\sum_{i=1}^n x_i \left(y_i - \frac{1}{n} \sum_{j=1}^n y_j \right) - a \sum_{i=1}^n x_i \left(x_i - \frac{1}{n} \sum_{j=1}^n x_j \right) \right] = \\ &= 2 \left[\sum_{i=1}^n x_i (y_i - \bar{y}) - a \sum_{i=1}^n x_i (x_i - \bar{x}) \right] = 0 \end{aligned}$$

$$\text{Per cui, } a = \frac{\sum_{i=1}^n x_i (y_i - \bar{y})}{\sum_{i=1}^n x_i (x_i - \bar{x})} = \frac{\sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i} = \frac{n(\sigma_{xy} + \bar{x}\bar{y}) - \bar{y}(n\bar{x})}{n(\sigma_x^2 + \bar{x}^2) - \bar{x}(n\bar{x})} = \frac{\sigma_{xy}}{\sigma_x^2} \text{ e}$$

$$b = \frac{1}{n} \sum_{i=1}^n \left(y_i - \frac{\sigma_{xy}}{\sigma_x^2} x_i \right)$$

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

$$\begin{cases} \frac{\partial f}{\partial a}(a, b) = -\sum_{i=1}^n 2x_i [y_i - (ax_i + b)] = 0 \\ \frac{\partial f}{\partial b}(a, b) = -\sum_{i=1}^n 2 [y_i - (ax_i + b)] = 0 \end{cases}$$

Dalla seconda equazione otteniamo $b = \frac{1}{n} \sum_{i=1}^n (y_i - ax_i)$. Sostituendo nella prima, abbiamo

$$\begin{aligned} \sum_{i=1}^n 2x_i \left[y_i - \left(ax_i + \frac{1}{n} \sum_{j=1}^n (y_j - ax_j) \right) \right] &= 2 \left[\sum_{i=1}^n x_i \left(y_i - \frac{1}{n} \sum_{j=1}^n y_j \right) - a \sum_{i=1}^n x_i \left(x_i - \frac{1}{n} \sum_{j=1}^n x_j \right) \right] = \\ &= 2 \left[\sum_{i=1}^n x_i (y_i - \bar{y}) - a \sum_{i=1}^n x_i (x_i - \bar{x}) \right] = 0 \end{aligned}$$

$$\text{Per cui, } a = \frac{\sum_{i=1}^n x_i (y_i - \bar{y})}{\sum_{i=1}^n x_i (x_i - \bar{x})} = \frac{\sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i} = \frac{n(\sigma_{xy} + \bar{x}\bar{y}) - \bar{y}(n\bar{x})}{n(\sigma_x^2 + \bar{x}^2) - \bar{x}(n\bar{x})} = \frac{\sigma_{xy}}{\sigma_x^2} \text{ e}$$

$$b = \frac{1}{n} \sum_{i=1}^n \left(y_i - \frac{\sigma_{xy}}{\sigma_x^2} x_i \right) = \bar{y} - \bar{x} \frac{\sigma_{xy}}{\sigma_x^2}.$$

Retta di regressione (Metodo dei minimi quadrati)

Metodo dei minimi quadrati

Minimizzare la funzione $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, tale che

$$f(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Calcoliamo dunque i punti di minimo di questa funzione. Occorre risolvere il seguente sistema:

$$\begin{cases} \frac{\partial f}{\partial a}(a, b) = -\sum_{i=1}^n 2x_i [y_i - (ax_i + b)] = 0 \\ \frac{\partial f}{\partial b}(a, b) = -\sum_{i=1}^n 2 [y_i - (ax_i + b)] = 0 \end{cases}$$

Dalla seconda equazione otteniamo $b = \frac{1}{n} \sum_{i=1}^n (y_i - ax_i)$. Sostituendo nella prima, abbiamo

$$\begin{aligned} \sum_{i=1}^n 2x_i \left[y_i - \left(ax_i + \frac{1}{n} \sum_{j=1}^n (y_j - ax_j) \right) \right] &= 2 \left[\sum_{i=1}^n x_i \left(y_i - \frac{1}{n} \sum_{j=1}^n y_j \right) - a \sum_{i=1}^n x_i \left(x_i - \frac{1}{n} \sum_{j=1}^n x_j \right) \right] = \\ &= 2 \left[\sum_{i=1}^n x_i (y_i - \bar{y}) - a \sum_{i=1}^n x_i (x_i - \bar{x}) \right] = 0 \end{aligned}$$

$$\text{Per cui, } a = \frac{\sum_{i=1}^n x_i (y_i - \bar{y})}{\sum_{i=1}^n x_i (x_i - \bar{x})} = \frac{\sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i} = \frac{n(\sigma_{xy} + \bar{x}\bar{y}) - \bar{y}(n\bar{x})}{n(\sigma_x^2 + \bar{x}^2) - \bar{x}(n\bar{x})} = \frac{\sigma_{xy}}{\sigma_x^2} \text{ e}$$

$$b = \frac{1}{n} \sum_{i=1}^n \left(y_i - \frac{\sigma_{xy}}{\sigma_x^2} x_i \right) = \bar{y} - \bar{x} \frac{\sigma_{xy}}{\sigma_x^2}.$$

Il punto (a, b) individuato è l'unico punto stazionario della funzione f . Poiché f è una funzione polinomiale quadratica sempre positiva, il punto (a, b) è necessariamente l'unico punto di minimo della funzione.

Retta di regressione (Metodo dei minimi quadrati)

Teorema

La retta di regressione corrispondente alle osservazioni $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, ha equazione

$$y = \hat{a}x + \hat{b} \text{ con } \hat{a} = \frac{\sigma_{xy}}{\sigma_x^2} \text{ e } \hat{b} = \bar{y} - \bar{x} \frac{\sigma_{xy}}{\sigma_x^2}.$$

Retta di regressione (Metodo dei minimi quadrati)

Teorema

La retta di regressione corrispondente alle osservazioni $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, ha equazione

$$y = \hat{a}x + \hat{b} \text{ con } \hat{a} = \frac{\sigma_{xy}}{\sigma_x^2} \text{ e } \hat{b} = \bar{y} - \bar{x} \frac{\sigma_{xy}}{\sigma_x^2}.$$

Definizione (Valori stimati, residui, valore previsto)

Sia $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ un insieme di n osservazioni e sia $y = \hat{a}x + \hat{b}$ la corrispondente retta di regressione.

- Chiamiamo **valori stimati**, i numeri

$$\hat{y}_i = \hat{a}x_i + \hat{b}.$$

Retta di regressione (Metodo dei minimi quadrati)

Teorema

La retta di regressione corrispondente alle osservazioni $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, ha equazione

$$y = \hat{a}x + \hat{b} \text{ con } \hat{a} = \frac{\sigma_{xy}}{\sigma_x^2} \text{ e } \hat{b} = \bar{y} - \bar{x} \frac{\sigma_{xy}}{\sigma_x^2}.$$

Definizione (Valori stimati, residui, valore previsto)

Sia $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ un insieme di n osservazioni e sia $y = \hat{a}x + \hat{b}$ la corrispondente retta di regressione.

- Chiamiamo **valori stimati**, i numeri

$$\hat{y}_i = \hat{a}x_i + \hat{b}.$$

- Chiamiamo **residui**, i numeri

$$r_i = y_i - \hat{y}_i.$$

Retta di regressione (Metodo dei minimi quadrati)

Teorema

La retta di regressione corrispondente alle osservazioni $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, ha equazione

$$y = \hat{a}x + \hat{b} \text{ con } \hat{a} = \frac{\sigma_{xy}}{\sigma_x^2} \text{ e } \hat{b} = \bar{y} - \bar{x} \frac{\sigma_{xy}}{\sigma_x^2}.$$

Definizione (Valori stimati, residui, valore previsto)

Sia $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ un insieme di n osservazioni e sia $y = \hat{a}x + \hat{b}$ la corrispondente retta di regressione.

- Chiamiamo **valori stimati**, i numeri

$$\hat{y}_i = \hat{a}x_i + \hat{b}.$$

- Chiamiamo **residui**, i numeri

$$r_i = y_i - \hat{y}_i.$$

- Dato un valore $x_0 \neq x_i$, per ogni $i = 1, \dots, n$, chiamiamo **valore previsto** di y_0 , il numero

$$\hat{y}_0 = \hat{a}x_0 + \hat{b}.$$

Esercizio sulla comparazione di variabili statistiche

Esercizio 1.14 (Bramanti)

Si considerino i seguenti dati, raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di "dati" e y il tempo, in secondo, impiegato dal computer per processarli.

x	y
211	112
332	155
322	131
435	208
275	138

Esercizio sulla comparazione di variabili statistiche

Esercizio 1.14 (Bramanti)

Si considerino i seguenti dati, raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di "dati" e y il tempo, in secondo, impiegato dal computer per processarli.

x	y
211	112
332	155
322	131
435	208
275	138

1. Si tracci uno scatterplot dei dati.

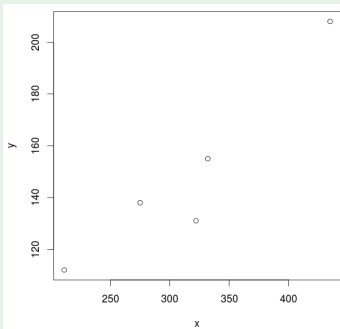
Esercizio sulla comparazione di variabili statistiche

Esercizio 1.14 (Bramanti)

Si considerino i seguenti dati, raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di "dati" e y il tempo, in secondo, impiegato dal computer per processarli.

x	y
211	112
332	155
322	131
435	208
275	138

1. Si tracci uno scatterplot dei dati.



Esercizio sulla comparazione di variabili statistiche

Esercizio 1.14 (Bramanti)

Si considerino i seguenti dati, raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di "dati" e y il tempo, in secondo, impiegato dal computer per processarli.

x	y
211	112
332	155
322	131
435	208
275	138

2. Si calcoli il coefficiente di correlazione tra le variabili.

Esercizio sulla comparazione di variabili statistiche

Esercizio 1.14 (Bramanti)

Si considerino i seguenti dati, raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di "dati" e y il tempo, in secondo, impiegato dal computer per processarli.

x	y
211	112
332	155
322	131
435	208
275	138

2. Si calcoli il coefficiente di correlazione tra le variabili.

Medie: $\bar{x} = \frac{211+332+322+435+275}{5} = 315$; $\bar{y} = \frac{112+155+131+208+138}{5} = 148.8$.

Varianze: $\sigma_x^2 = \frac{211^2+332^2+322^2+435^2+275^2}{5} - 315^2 = 5430.8$;
 $\sigma_y^2 = \frac{112^2+155^2+131^2+208^2+138^2}{5} - 148.8^2 = 1332.7$.

Esercizio sulla comparazione di variabili statistiche

Esercizio 1.14 (Bramanti)

Si considerino i seguenti dati, raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di "dati" e y il tempo, in secondo, impiegato dal computer per processarli.

x	y
211	112
332	155
322	131
435	208
275	138

2. Si calcoli il coefficiente di correlazione tra le variabili.

Medie: $\bar{x} = \frac{211+332+322+435+275}{5} = 315$; $\bar{y} = \frac{112+155+131+208+138}{5} = 148.8$.

Varianze: $\sigma_x^2 = \frac{211^2+332^2+322^2+435^2+275^2}{5} - 315^2 = 5430.8$;

$\sigma_y^2 = \frac{112^2+155^2+131^2+208^2+138^2}{5} - 148.8^2 = 1332.7$.

Deviazioni standard: $\sigma_x = \sqrt{\sigma_x^2} = \sqrt{5430.8} \simeq 73.69396$; $\sigma_y = \sqrt{\sigma_y^2} = \sqrt{1332.7} \simeq 36.50616$.

Esercizio sulla comparazione di variabili statistiche

Esercizio 1.14 (Bramanti)

Si considerino i seguenti dati, raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di "dati" e y il tempo, in secondo, impiegato dal computer per processarli.

x	y
211	112
332	155
322	131
435	208
275	138

2. Si calcoli il coefficiente di correlazione tra le variabili.

Medie: $\bar{x} = \frac{211+332+322+435+275}{5} = 315$; $\bar{y} = \frac{112+155+131+208+138}{5} = 148.8$.

Varianze: $\sigma_x^2 = \frac{211^2+332^2+322^2+435^2+275^2}{5} - 315^2 = 5430.8$;

$\sigma_y^2 = \frac{112^2+155^2+131^2+208^2+138^2}{5} - 148.8^2 = 1332.7$.

Deviazioni standard: $\sigma_x = \sqrt{\sigma_x^2} = \sqrt{5430.8} \simeq 73.69396$; $\sigma_y = \sqrt{\sigma_y^2} = \sqrt{1332.7} \simeq 36.50616$.

Covarianza: $\sigma_{xy} = \frac{211 \cdot 112 + 332 \cdot 155 + 322 \cdot 131 + 435 \cdot 208 + 275 \cdot 138}{5} - 315 \cdot 148.8 = 2268.8$.

Esercizio sulla comparazione di variabili statistiche

Esercizio 1.14 (Bramanti)

Si considerino i seguenti dati, raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di "dati" e y il tempo, in secondo, impiegato dal computer per processarli.

x	y
211	112
332	155
322	131
435	208
275	138

2. Si calcoli il coefficiente di correlazione tra le variabili.

Medie: $\bar{x} = \frac{211+332+322+435+275}{5} = 315$; $\bar{y} = \frac{112+155+131+208+138}{5} = 148.8$.

Varianze: $\sigma_x^2 = \frac{211^2+332^2+322^2+435^2+275^2}{5} - 315^2 = 5430.8$;

$\sigma_y^2 = \frac{112^2+155^2+131^2+208^2+138^2}{5} - 148.8^2 = 1332.7$.

Deviazioni standard: $\sigma_x = \sqrt{\sigma_x^2} = \sqrt{5430.8} \simeq 73.69396$; $\sigma_y = \sqrt{\sigma_y^2} = \sqrt{1332.7} \simeq 36.50616$.

Covarianza: $\sigma_{xy} = \frac{211 \cdot 112 + 332 \cdot 155 + 322 \cdot 131 + 435 \cdot 208 + 275 \cdot 138}{5} - 315 \cdot 148.8 = 2268.8$.

Coefficiente di correlazione: $\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{2268.8}{73.69396 \cdot 36.50616} \simeq 0.843$.

Esercizio sulla comparazione di variabili statistiche

Esercizio 1.14 (Bramanti)

Si considerino i seguenti dati, raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di "dati" e y il tempo, in secondo, impiegato dal computer per processarli.

x	y
211	112
332	155
322	131
435	208
275	138

2. Si calcoli il coefficiente di correlazione tra le variabili.

Medie: $\bar{x} = \frac{211+332+322+435+275}{5} = 315$; $\bar{y} = \frac{112+155+131+208+138}{5} = 148.8$.

Varianze: $\sigma_x^2 = \frac{211^2+332^2+322^2+435^2+275^2}{5} - 315^2 = 5430.8$;

$\sigma_y^2 = \frac{112^2+155^2+131^2+208^2+138^2}{5} - 148.8^2 = 1332.7$.

Deviazioni standard: $\sigma_x = \sqrt{\sigma_x^2} = \sqrt{5430.8} \simeq 73.69396$; $\sigma_y = \sqrt{\sigma_y^2} = \sqrt{1332.7} \simeq 36.50616$.

Covarianza: $\sigma_{xy} = \frac{211 \cdot 112 + 332 \cdot 155 + 322 \cdot 131 + 435 \cdot 208 + 275 \cdot 138}{5} - 315 \cdot 148.8 = 2268.8$.

Coefficiente di correlazione: $\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{2268.8}{73.69396 \cdot 36.50616} \simeq 0.843$.

Pertanto tra i dati esiste una **forte correlazione diretta** ($\rho \simeq 1$).

Esercizio sulla comparazione di variabili statistiche

Esercizio 1.14 (Bramanti)

Si considerino i seguenti dati, raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di "dati" e y il tempo, in secondo, impiegato dal computer per processarli.

x	y
211	112
332	155
322	131
435	208
275	138

3. Si scriva l'equazione della retta di regressione e la si disegni sullo scatterplot.

Esercizio sulla comparazione di variabili statistiche

Esercizio 1.14 (Bramanti)

Si considerino i seguenti dati, raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di "dati" e y il tempo, in secondo, impiegato dal computer per processarli.

x	y
211	112
332	155
322	131
435	208
275	138

3. Si scriva l'equazione della retta di regressione e la si disegni sullo scatterplot.

Retta di regressione: $y = \hat{a}x + \hat{b}$, dove $\hat{a} = \frac{\sigma_{xy}}{\sigma_x^2}$ e $\hat{b} = \bar{y} - \bar{x}\hat{a}$.

Esercizio sulla comparazione di variabili statistiche

Esercizio 1.14 (Bramanti)

Si considerino i seguenti dati, raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di "dati" e y il tempo, in secondo, impiegato dal computer per processarli.

x	y
211	112
332	155
322	131
435	208
275	138

3. Si scriva l'equazione della retta di regressione e la si disegni sullo scatterplot.

Retta di regressione: $y = \hat{a}x + \hat{b}$, dove $\hat{a} = \frac{\sigma_{xy}}{\sigma_x^2}$ e $\hat{b} = \bar{y} - \bar{x}\hat{a}$.

Dunque $\hat{a} = \frac{2268.8}{5430.8} \simeq 0.42$ e $\hat{b} = 148.8 - 315 \cdot 0.42 \simeq 17.20$.

Esercizio sulla comparazione di variabili statistiche

Esercizio 1.14 (Bramanti)

Si considerino i seguenti dati, raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di "dati" e y il tempo, in secondo, impiegato dal computer per processarli.

x	y
211	112
332	155
322	131
435	208
275	138

3. Si scriva l'equazione della retta di regressione e la si disegni sullo scatterplot.

Retta di regressione: $y = \hat{a}x + \hat{b}$, dove $\hat{a} = \frac{\sigma_{xy}}{\sigma_x^2}$ e $\hat{b} = \bar{y} - \bar{x}\hat{a}$.

Dunque $\hat{a} = \frac{2268.8}{5430.8} \simeq 0.42$ e $\hat{b} = 148.8 - 315 \cdot 0.42 \simeq 17.20$.

Pertanto, $y = 0.42x + 17.20$.

Esercizio sulla comparazione di variabili statistiche

Esercizio 1.14 (Bramanti)

Si considerino i seguenti dati, raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di "dati" e y il tempo, in secondo, impiegato dal computer per processarli.

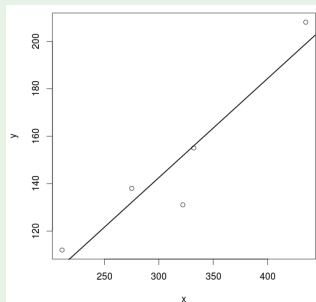
x	y
211	112
332	155
322	131
435	208
275	138

3. Si scriva l'equazione della retta di regressione e la si disegni sullo scatterplot.

Retta di regressione: $y = \hat{a}x + \hat{b}$, dove $\hat{a} = \frac{\sigma_{xy}}{\sigma_x^2}$ e $\hat{b} = \bar{y} - \bar{x}\hat{a}$.

Dunque $\hat{a} = \frac{2268.8}{5430.8} \simeq 0.42$ e $\hat{b} = 148.8 - 315 \cdot 0.42 \simeq 17.20$.

Pertanto, $y = 0.42x + 17.20$.



Esercizio sulla comparazione di variabili statistiche

Esercizio 1.14 (Bramanti)

Si considerino i seguenti dati, raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di "dati" e y il tempo, in secondo, impiegato dal computer per processarli.

x	y
211	112
332	155
322	131
435	208
275	138

4. Si calcoli il tempo previsto per processare 200, 300, 400 e 500 dati.

Esercizio sulla comparazione di variabili statistiche

Esercizio 1.14 (Bramanti)

Si considerino i seguenti dati, raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di "dati" e y il tempo, in secondo, impiegato dal computer per processarli.

x	y
211	112
332	155
322	131
435	208
275	138

4. Si calcoli il tempo previsto per processare 200, 300, 400 e 500 dati.

Attraverso la retta di regressione $y = 0.42x + 17.20$ possiamo fare le seguenti previsioni:

Esercizio sulla comparazione di variabili statistiche

Esercizio 1.14 (Bramanti)

Si considerino i seguenti dati, raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di "dati" e y il tempo, in secondo, impiegato dal computer per processarli.

x	y
211	112
332	155
322	131
435	208
275	138

4. Si calcoli il tempo previsto per processare 200, 300, 400 e 500 dati.

Attraverso la retta di regressione $y = 0.42x + 17.20$ possiamo fare le seguenti previsioni:

- per $x = 200$, $\hat{y} = 0.42 \cdot 200 + 17.20 \simeq 101$;

Esercizio sulla comparazione di variabili statistiche

Esercizio 1.14 (Bramanti)

Si considerino i seguenti dati, raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di "dati" e y il tempo, in secondo, impiegato dal computer per processarli.

x	y
211	112
332	155
322	131
435	208
275	138

4. Si calcoli il tempo previsto per processare 200, 300, 400 e 500 dati.

Attraverso la retta di regressione $y = 0.42x + 17.20$ possiamo fare le seguenti previsioni:

- per $x = 200$, $\hat{y} = 0.42 \cdot 200 + 17.20 \simeq 101$;
- per $x = 300$, $\hat{y} = 0.42 \cdot 300 + 17.20 \simeq 143$;

Esercizio sulla comparazione di variabili statistiche

Esercizio 1.14 (Bramanti)

Si considerino i seguenti dati, raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di "dati" e y il tempo, in secondo, impiegato dal computer per processarli.

x	y
211	112
332	155
322	131
435	208
275	138

4. Si calcoli il tempo previsto per processare 200, 300, 400 e 500 dati.

Attraverso la retta di regressione $y = 0.42x + 17.20$ possiamo fare le seguenti previsioni:

- per $x = 200$, $\hat{y} = 0.42 \cdot 200 + 17.20 \simeq 101$;
- per $x = 300$, $\hat{y} = 0.42 \cdot 300 + 17.20 \simeq 143$;
- per $x = 400$, $\hat{y} = 0.42 \cdot 400 + 17.20 \simeq 184$;

Esercizio sulla comparazione di variabili statistiche

Esercizio 1.14 (Bramanti)

Si considerino i seguenti dati, raccolti per stimare il tempo che impiega un computer a processare dati: x è il numero di "dati" e y il tempo, in secondo, impiegato dal computer per processarli.

x	y
211	112
332	155
322	131
435	208
275	138

4. Si calcoli il tempo previsto per processare 200, 300, 400 e 500 dati.

Attraverso la retta di regressione $y = 0.42x + 17.20$ possiamo fare le seguenti previsioni:

- per $x = 200$, $\hat{y} = 0.42 \cdot 200 + 17.20 \simeq 101$;
- per $x = 300$, $\hat{y} = 0.42 \cdot 300 + 17.20 \simeq 143$;
- per $x = 400$, $\hat{y} = 0.42 \cdot 400 + 17.20 \simeq 184$;
- per $x = 500$, $\hat{y} = 0.42 \cdot 500 + 17.20 \simeq 226$.