Caso studio 12

Per studiare la curva di domanda di un bene che sta per essere introdotto sul mercato, si rilevano i dati riguardanti il prezzo imposto e il numero di pezzi venduti in 7 punti vendita pilota, nell'arco di una settimana. I dati sono riportati in tabella:

Punto vendita	Prezzo (in euro)	Numero di pezzi
1	12,6	47
2	14,5	42
3	12	54
4	11,6	57
5	12,9	51
6	13,5	46
7	13,9	39

Come varia la quantità venduta in funzione del prezzo? Come prevedere la quantità venduta per un certo valore del prezzo?

Regressione

- Concetto utilizzato quando *entrambi i caratteri* (X e Y) *sono quantitativi*
- In questo caso si preferisce utilizzare direttamente la distribuzione doppia unitaria

X	Y
x_1	y_1
x_2	y_2
…	…
x_n	y_n

- Solitamente si intende che la variabile Y dipenda da X. Quindi X è chiamata ***variabile indipendente*** (o ***esplicativa***) mentre Y è chiamata ***variabile dipendente*** (o ***risposta***).

- L'obiettivo è capire come X influenza Y e approssimare tale relazione tramite una semplice funzione matematica

$$y = f(x)$$

- Un'analisi preliminare può essere effettuata tramite un ***grafico a dispersione*** che consiste nel rappresentare punti di coordinate

$$(x_i, y_i), \quad i = 1, \cdots, n$$

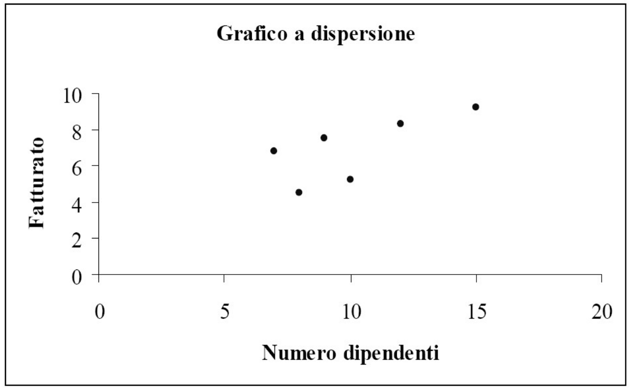
Esempio

- Per un collettivo di n = 6 imprese sono state rilevate le modalità dei caratteri *numero di dipendenti* (X) e *fatturato* (Y).

Dipendenti	Fatturato
15	9,2
10	5,2
8	4,5
7	6,8
12	8,3
9	7,5

- Sulla base di questi dati si vuole stabilire come il numero di dipendenti *influenzi* il fatturato di un'impresa.

- Dal *grafico a dispersione* si nota che al crescere di X cresce Y.

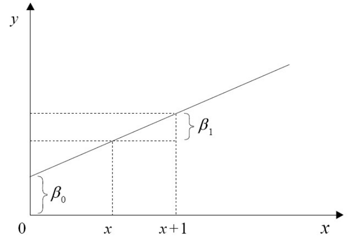


Regressione lineare semplice

- In questo ambito, per approssimare la relazione tra Y e X si utilizza una funzione lineare

$$Y = \beta_0 + \beta_1 X$$

dove β_0 (*intercetta*) e β_1 (*coefficiente angolare*) sono chiamati ***parametri della retta interpolatrice***.



- I parametri della retta (β_0 e β_1) vanno scelti in modo da approssimare al meglio la relazione tra Y e X, ossia *minimizzando l'errore di previsione*. Il metodo normalmente utilizzato a tale fine è conosciuto con il nome di *metodo dei minimi quadrati*.

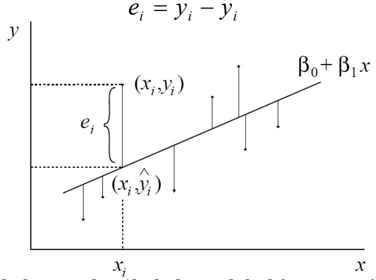
Metodo dei minimi quadrati

- Per un certo valore dei parametri (β_0 e β_1) si definisce ***valore teorico*** (o ***previsto***) corrispondente alla *i*-esima osservazione

$$\hat{y}_i = \beta_0 + \beta_1 x_i$$

- Il corrispondente ***errore di previsione*** (o ***residuo***) è

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



- La ***somma dei quadrati dei residui*** è una misura complessiva dell'errore di previsione, definita come

$$G(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

- Il ***metodo dei minimi quadrati*** consiste nel *minimizzare la somma dei quadrati dei residui, G*(β_0,β_1), rispetto ai parametri β_0 e β_1

- Si può dimostrare che il minimo si raggiunge quando il *coefficiente angolare* e l'*intercetta* sono pari, rispettivamente, a

$$\hat{\beta}_1 = \frac{\sigma_{XY}}{\sigma_X^2} \quad e \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

- La ***retta di regressione*** di Y da X si ottiene sostituendo a β_0 e β_1 i corrispondenti valori trovati con il metodo dei minimi quadrati

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

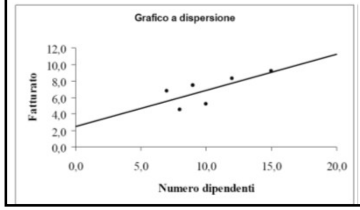
Esempio

- Per il collettivo di n = 6 imprese si ha:

x_i	y_i	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
15	9,2	11,036	23,361
10	5,2	0,286	0,028
8	4,5	5,236	4,694
7	6,8	0,369	10,028
12	8,3	2,536	3,361
9	7,5	-0,681	1,361
61,0	41,5	18,783	42,833

$$\bar{x} = 61/6 = 10{,}17$$

$$\bar{y} = 41{,}5/6 = 6{,}92$$



$$\sigma_X^2 = 42{,}833/6 = 7{,}139$$

$$\sigma_{XY} = 18{,}783/6 = 3{,}13$$

$$\hat{\beta}_1 = \frac{18{,}783/6}{42{,}833/6} = 0{,}439$$

$$\hat{\beta}_0 = 6{,}917 - 0{,}439 \cdot 10{,}167 = 2{,}458$$

- Il calcolo della varianza e covarianza può essere effettuato con le formule semplificate nel modo seguente

x_i	y_i	$x_i y_i$	x_i^2
15	9,2	138,0	225,0
10	5,2	52,0	100,0
8	4,5	36,0	64,0
7	6,8	47,6	49,0
12	8,3	99,6	144,0
9	7,5	67,5	81,0
61,0	41,5	440,7	663,0

$$\sigma_{XY} = 440{,}7/6 - 10{,}167 \cdot 6{,}917 = 3{,}13$$

$$\sigma_X^2 = 663/6 - 10{,}167^2 = 7{,}139$$

Esercizio

Si considerino i dati del Caso Studio 12.

Si disegni il grafico a dispersione della quantità rispetto al prezzo e si commenti il risultato.

Si calcolino i parametri della retta di regressione con il metodo dei minimi quadrati e si determini la funzione della curva di domanda.

Si interpreti il valore del coefficiente angolare.

Indice di determinazione

- Per misurare la bontà dell'adattamento della retta ai punti si può utilizzare la ***varianza residua*** data da

$$SQE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

- La devianza residua è sempre non negativa e quanto più è vicina a 0, tanto migliore è la bontà dell'adattamento della retta ai punti.
- Il massimo della varianza residua è dato dalla ***varianza totale*** di Y,

$$\sigma_Y^2 = SQT = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2$$

- La differenza tra la varianza totale e quella residua è chiamata ***varianza spiegata*** che è sempre non negativa

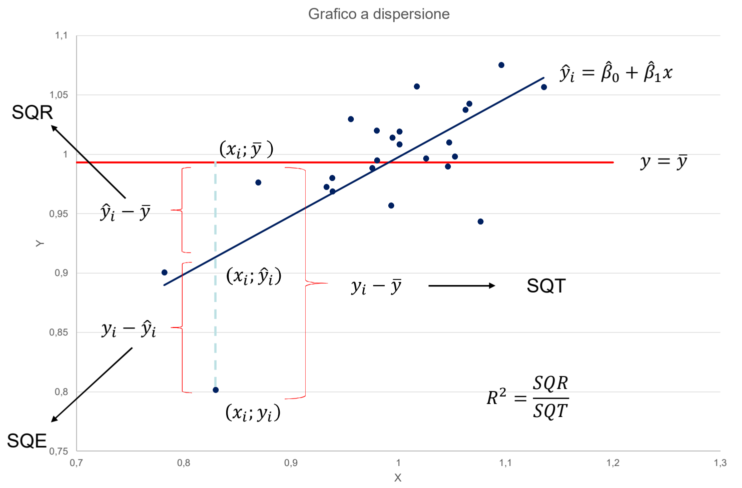
$$SQR = SQT - SQE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

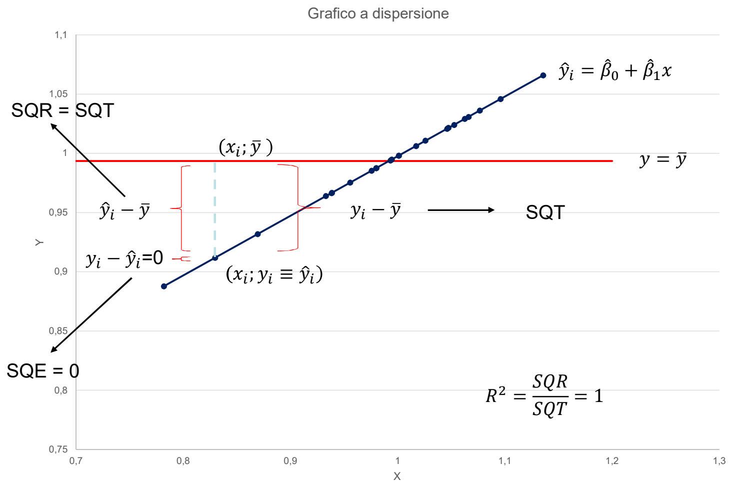
- E' quindi possibile costruire un *indice relativo* (varia tra 0 e 1) per misurare la bontà dell'adattamento, chiamato di ***indice di determinazione***, come

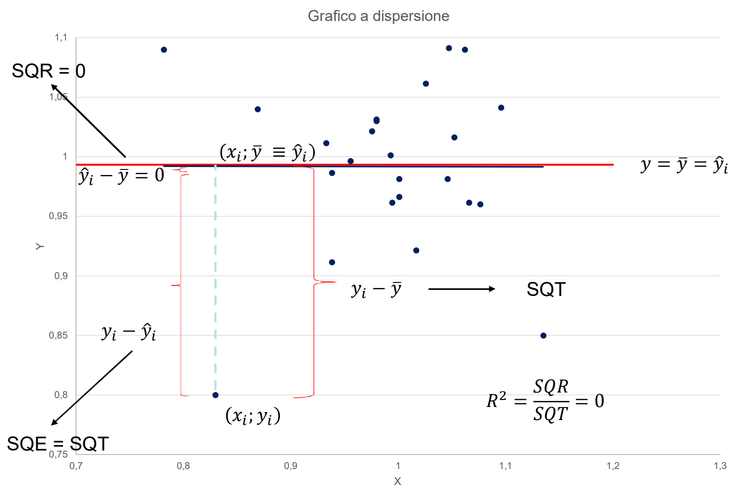
$$R^2 = 1 - \frac{SQE}{SQT} = \frac{SQR}{SQT}$$

- Una *formula alternativa* per l'indice di determinazione è

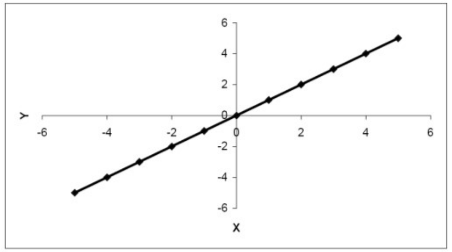
$$R^2 = \frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2}$$



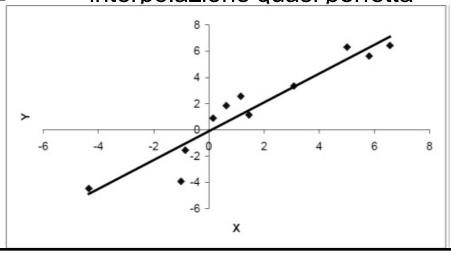




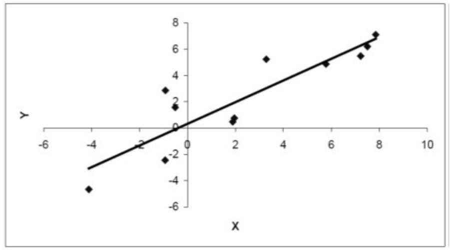
- $R^2 = 1 \;\Rightarrow$ interpolazione perfetta



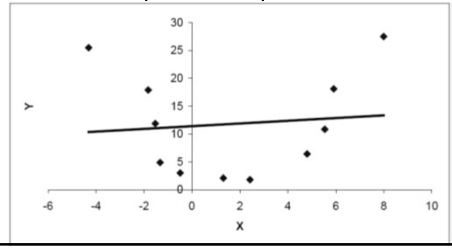
- $R^2 = 0{,}92$ interpolazione quasi perfetta



- $R^2 = 0{,}72$ interpolazione buona



- $R^2 = 0{,}02$ interpolazione pessima



Esempio

- Per il collettivo di *n* = 6 imprese si ha:

x_i	y_i	$\hat{y}_i$	$(y_i - \hat{y}_i)^2$	y_i^2
15	9,2	9,0362	0,0268	84,64
10	5,2	6,8436	2,7014	27,04
8	4,5	5,9665	2,1507	20,25
7	6,8	5,5280	1,6179	46,24
12	8,3	7,7206	0,3357	68,89
9	7,5	6,4051	1,1989	56,25
61,0	41,5	41,5	8,0314	303,31

$$\sigma_Y^2 = SQT = 303{,}31/6 - 6{,}917^2 = 2{,}707$$

$$SQE = 8{,}031/6 = 1{,}339$$

$$R^2 = 1 - \frac{1{,}339}{2{,}707} = 0{,}506$$

Esercizio

Si considerino i dati del Caso Studio 12.

Si calcoli il valore dell'indice di determinazione e se ne interpreti il risultato.

Qual è il valore dell'indice di correlazione di Bravais-Pearson per i dati a disposizione?

Previsione

- Una volta calcolati i parametri della retta di regressione (β_0,β_1), è possibile ***prevedere il valore di*** *Y* in corrispondenza di un nuovo valore di *X* indicato con x_{n+1}. Il valore previsto si calcola come:

$$\hat{y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{n+1}$$

- La previsione è particolarmente utilizzata nell'analisi di serie storiche quando, ad esempio, si vuole prevedere il *prodotto interno lordo* di un paese in base ai valori osservati di questo indicatore in un certo periodo di tempo.

Esempio

- Sulla base dei dati dell'esempio precedente si vuole prevedere il fatturo per un'impresa che ha 20 dipendenti. Si ha:

$$\hat{y}_7 = 2{,}458 + 0{,}439 \cdot 20 = 11{,}229$$

Esercizio

Si considerino i dati del Caso Studio 12.

Si preveda la quantità venduta in corrispondenza di un prezzo di 14 euro.

Ha senso utilizzare i dati a disposizione per prevedere la quantità venduta in corrispondenza di un prezzo di 7 euro?

Dove e come studiare

- Libro di testo: S. Borra, A. Di Ciaccio (2014), Cap. 16 (escluso paragrafo 16.6)
- Svolgere 'Esercitazione 4', esercizi 1 e 2.
- Svolgere gli esercizi nel file 'Esercizi di statistica bivariata.xls', fogli non precedentemente svolti.