

Relazioni tra variabili (fenomeni) aziendali

- giornate di assenza dal lavoro è *collegato* a qualifica professionale, anzianità, sesso, ...dei dipendenti ?
- incidenti sul lavoro *sono collegati* a orario di lavoro giornaliero e/o settimanale dei lavoratori?
- le quantità prodotte *sono collegate* ai costi degli input di produzione ?
- prezzo di un prodotto è *collegato* alle quantità e caratteristiche delle sue componenti ?
- vendite di un prodotto *sono collegate* al prezzo del prodotto e alla spese di promozione ?
- vendite di un prodotto *sono collegate* allo disponibilità di spazio espositivo del negozio, al prezzo del prodotto e alla spese di promozione ?

Analisi di regressione (modello di regressione lineare)

- obiettivo principale: **investigare** su eventuali **relazioni empiriche tra variabili** per verificare **se e come** una variabile (**dipendente, Y**) varia al variare delle altre variabili (**esplicative o indipendenti, X**), individuando una opportuna **funzione analitica** che esprima tale relazione.
- caratterizzata da **semplicità** intrinseca dei **modelli** utilizzati, basati essenzialmente su **funzioni lineari**
- anche se non tutte le relazioni funzionali possono essere espresse attraverso modelli lineari, una **prima analisi** fondata su **forme funzionali semplici** è comunque un **buon punto di partenza** per passare poi ad eventuali modelli più complessi
- Una sola X: regressione semplice
- Due o più X: regressione multipla

Utilizzi analisi di regressione

Descrizione: si vuole rappresentare tramite una funzione l'andamento dei valori d'una variabile al variare dell'altra.

Interpretazione: si cerca di mettere in evidenza i nessi causali fra le variabili, per confermare (o smentire) una teoria (economica nel ns. caso).

Previsione: si tenta di valutare in maniera attendibile il valore che assumerà la variabile dipendente in corrispondenza d'un valore noto della variabile esplicativa (o delle variabili esplicative, nel caso di regressione multipla).

In generale, meglio dire: *associazioni*

Analisi di regressione semplice

Presupposto:

- esiste una variabile X (esplicativa) che “agisce” su una variabile Y (dipendente, risposta, risultato)

La scelta del ruolo delle variabili (quale dipendente e quali esplicative) è una scelta extra-statistica

Esempio vendite e superficie supermercato

Rilevazione delle variabili su 10 negozi

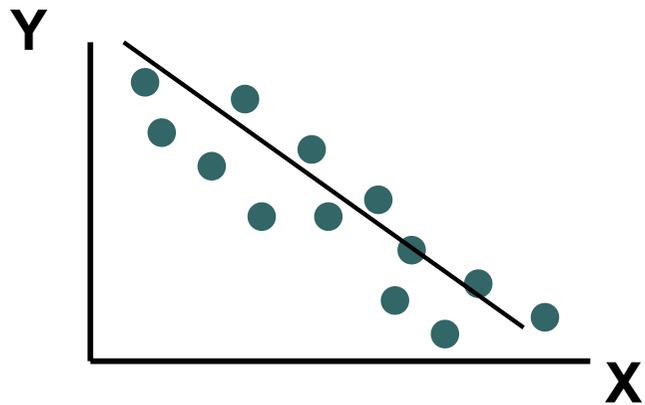
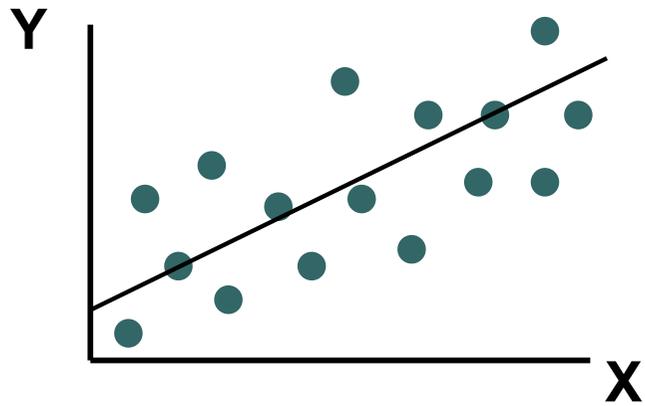
Volume vendite settimanali (*100 €)	Spazio espositivo (m ²)
43,2	95
132,0	144
155,0	210
76,0	156
100,9	188
187,4	321
185,0	250
60,7	115
82,9	178
61,3	105

Per decisioni su un nuovo punto vendita, per una catena di supermercati può essere di interesse studiare la relazione tra dimensione del negozio e vendite settimanali

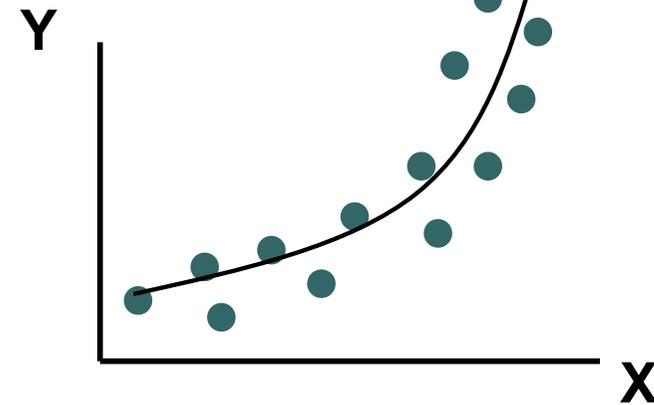
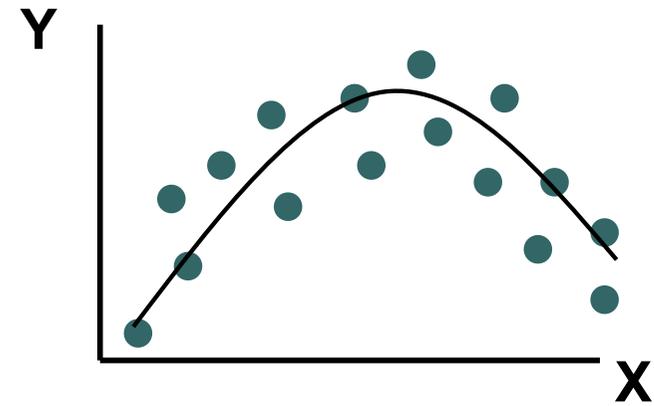
Diagramma di dispersione (*scatter plot*) per esame grafico della relazione tra le variabili

Analisi grafica: esempi di relazione /1

Relazione Lineare

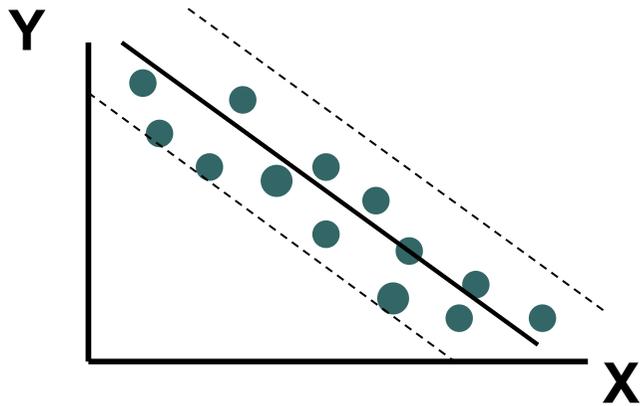
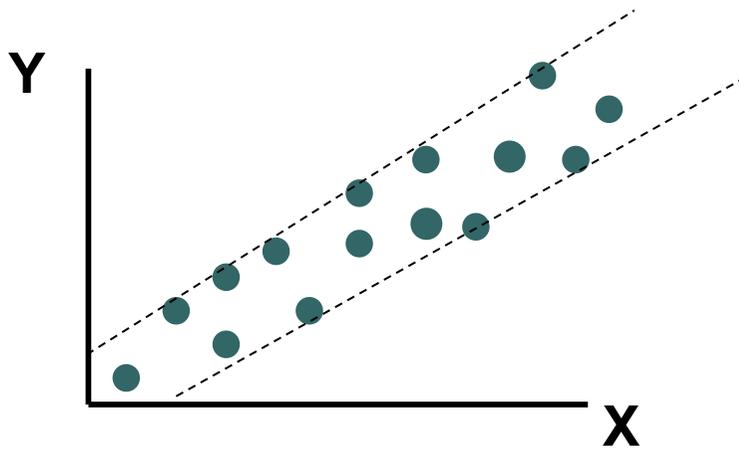


Relazione non lineare

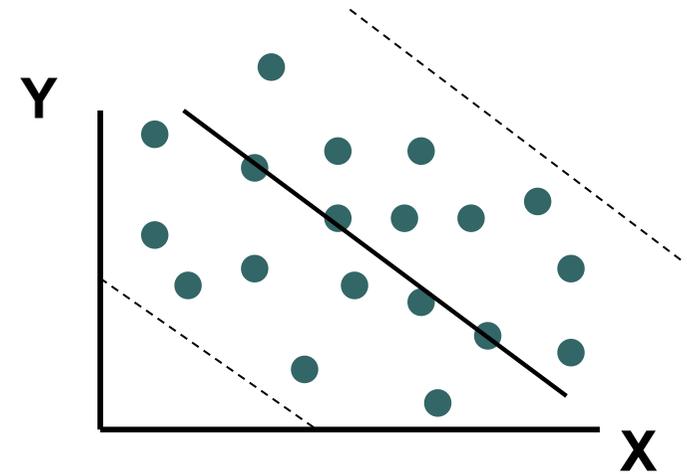
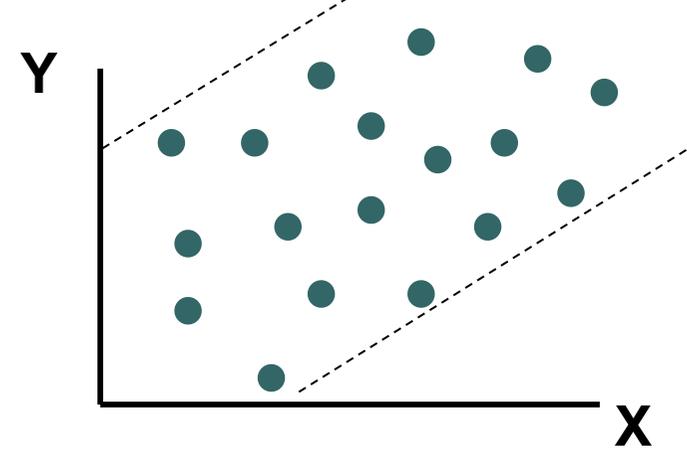


Analisi grafica: esempi di relazione /2

Relazione forte

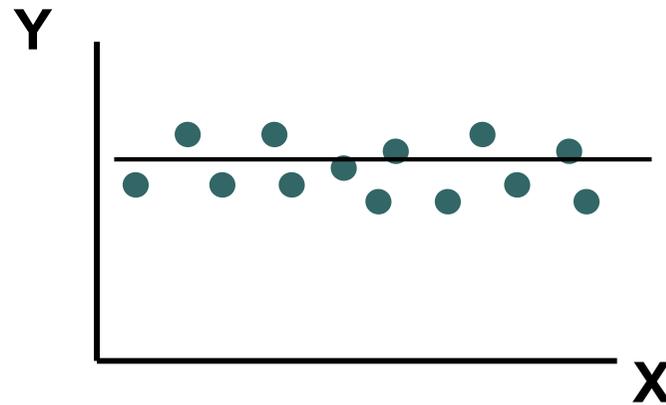
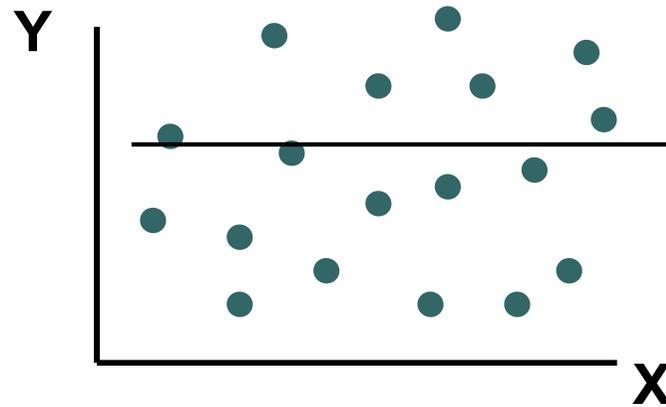


Relazione debole

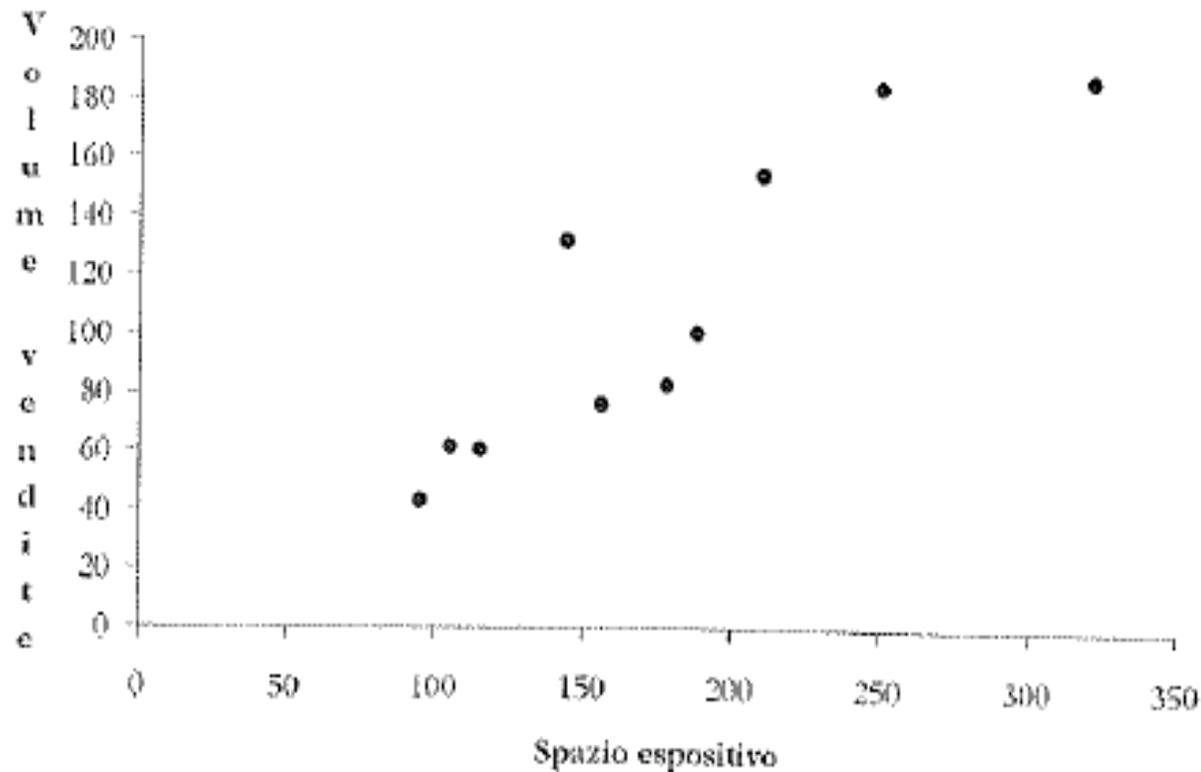


Analisi grafica: esempi di relazione /3

Nessuna relazione



Analisi grafica dati supermercati



E' ragionevole ipotizzare che a maggiori spazi espositivi tendano a corrispondere valori più elevati delle vendite:

Relazione espressa da un modello lineare (retta di regressione) dove:

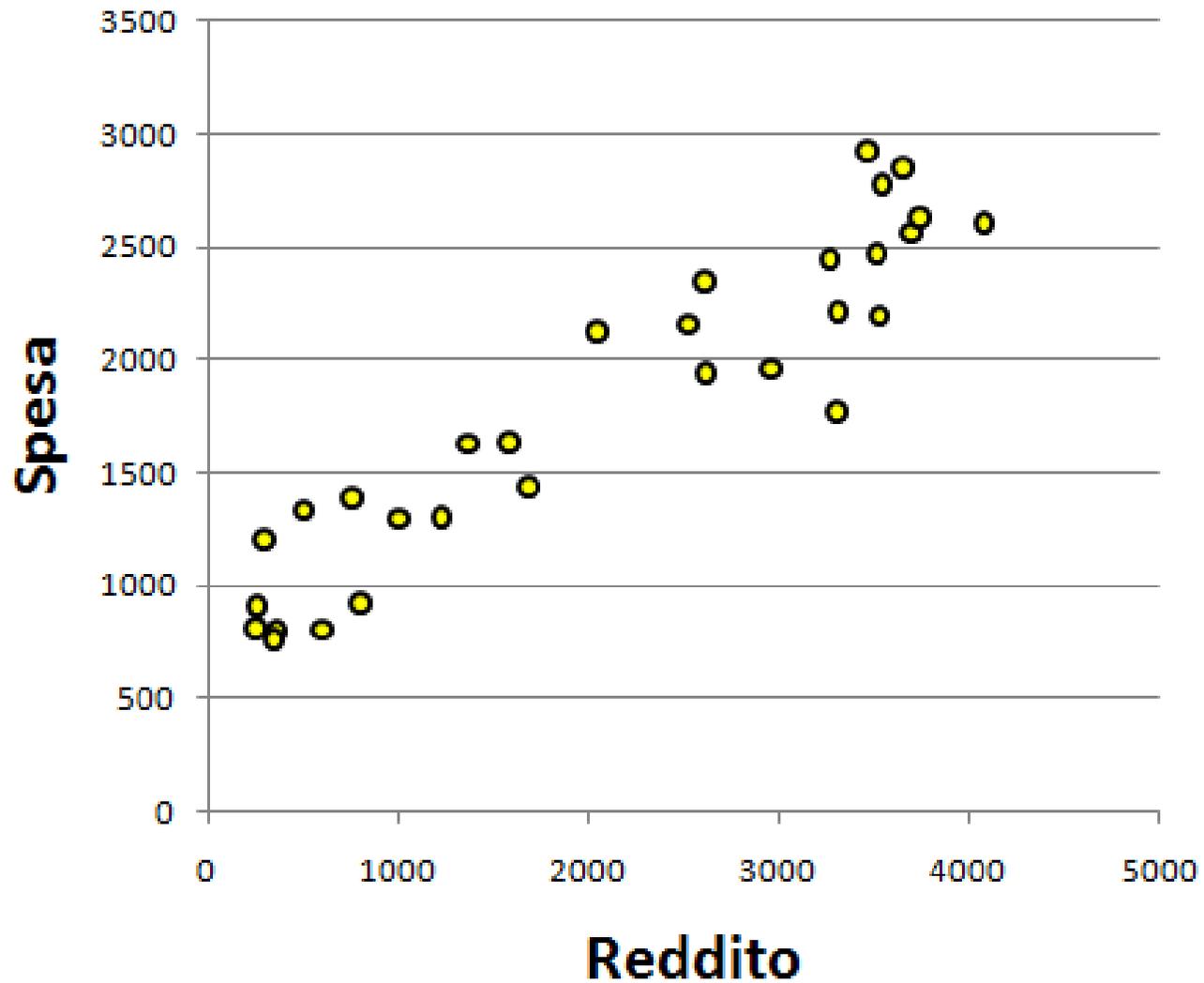
spazio espositivo

→ var. esplicativa (X)

e volume delle vendite

→ var. risposta (Y)

Analisi grafica dati su reddito e spesa mensile



Modello di regressione lineare semplice

- campione di n unità statistiche, osservati i valori relativi a **due distinte variabili**:

Y variabile dipendente o variabile risposta

X variabile indipendente o variabile esplicativa

Sulla base dei **dati osservati** e di **alcune assunzioni** può essere formulata la seguente relazione lineare

$$Y_i = \alpha + \beta X_i + u_i \quad i = 1, 2, \dots, n$$

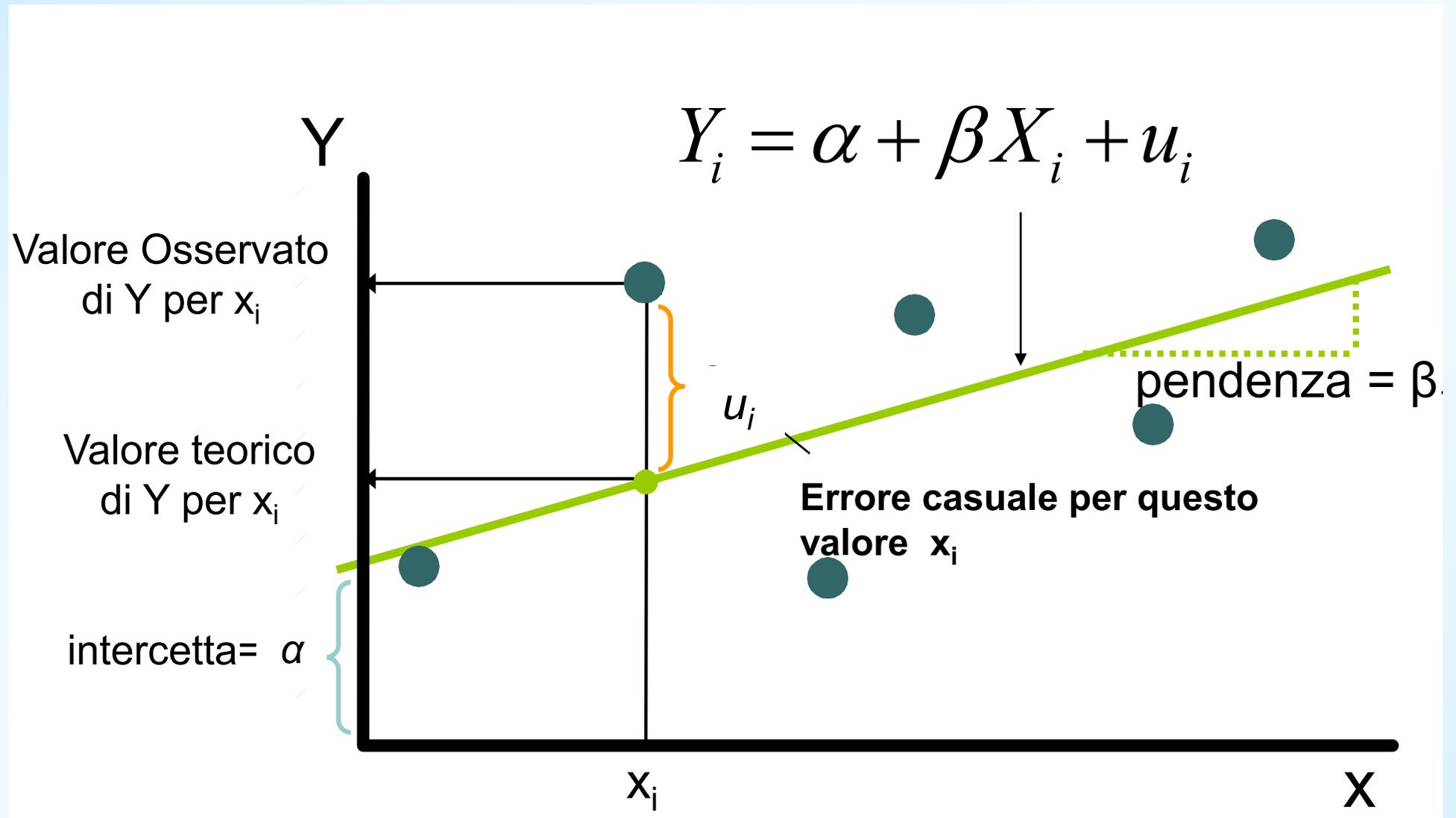
con

α e β sono costanti, parametri del modello di regressione

α intercetta e β coefficiente di regressione

u_i variabile casuale - detta errore - che si ipotizza distribuita normalmente con media $E(u) = 0$ e varianza $E(u^2) = \sigma^2_u$

Modello di regressione lineare semplice



Si assume che ciascun **valore osservato** $Y_i = f(X_i)$ con f funzione lineare + termine di disturbo u_i (*relazione non esatta*)

Modello di regressione lineare semplice

u_i termine di disturbo: incapacità del modello di riprodurre i dati osservati (relazione *statistica* e non esatta)

E' il risultato di:

1. Errori e carenze nella misurazione e nella rilevazione di "Y" e di "x"
2. Inadeguatezza della "semplice" relazione lineare
3. Insufficienza della variabile x a "spiegare" da sola la Y

Come determinare (stimare) l'equazione della retta (parametri α e β) relativa ai dati osservati ?