Analisi di regressione

L'analisi di regressione è una tecnica statistica utilizzata per determinare una relazione tra una variabile dipendente e un insieme di fattori esplicativi. La variabile dipendente, indicata come variabile y, è il valore che stiamo cercando di determinare in base ai fattori esplicativi. I fattori esplicativi, indicati come variabili x, vengono anche chiamati fattori indipendenti, variabili predittive o semplicemente fattori modello. L'analisi di regressione aiuta gli analisti a scoprire la sensibilità della variabile dipendente ai cambiamenti nei fattori esplicativi. Queste sensibilità sono essenziali per una corretta gestione del rischio. Esistono tre tipi di dati comunemente utilizzati nell'analisi di regressione: serie temporali, sezioni trasversali e dati raggruppati.

* Serie temporali: dati raccolti in un periodo di tempo. Nelle serie economiche e finanziarie questi dati si riferiscono spesso a rendimenti di mercato, rendimenti dell'indice, prezzi e valori delle attività, PIL, disoccupazione, tassi di interesse, ecc. Questi dati vengono raccolti a intervalli di tempo uguali come giornaliero, mensile, trimestrale, ecc.
* Sezione trasversale: dati raccolti per una famiglia di variabili nello stesso momento. Ad esempio, nell'analisi fondamentale raccogliamo spesso informazioni specifiche dell'azienda come il rapporto prezzo / utili, il rapporto prezzo / valore contabile, il rapporto debito / capitale netto o il volume medio giornaliero degli scambi.
* Dati raggruppati: dati che sono una combinazione di serie temporali e dati cross section.

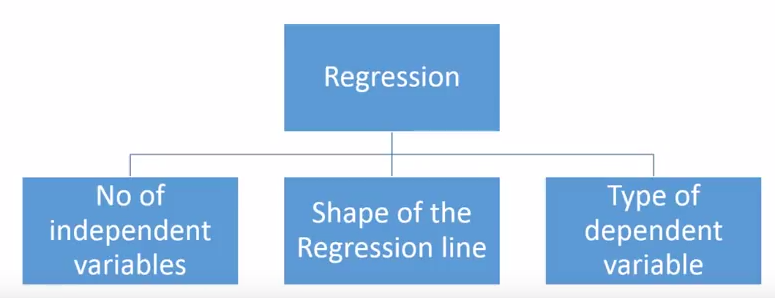
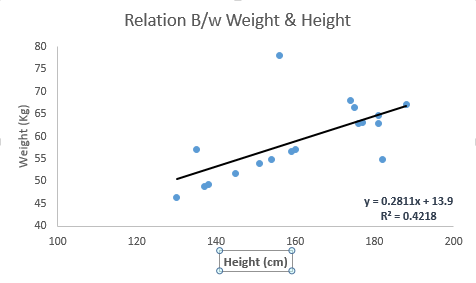
Regressione lineare

Un modello di regressione lineare è un modello in cui formuliamo una relazione lineare tra la variabile dipendente e il / i fattore / i esplicativo / i.

Nel caso in cui abbiamo un singolo fattore esplicativo, l'analisi è chiamata un semplice modello di regressione e ha la forma:

Y = b 0+ b 1X + ε

dove y è la variabile dipendente (quello che stiamo cercando di prevedere), x è il fattore esplicativo (quello che stiamo usando per predire), e ε è il rumore casuale(errore). Inoltre, la variabile dipendente y, i fattori esplicativi x e il termine di errore ε sono vettori di colonne di valori.

[](https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Regression_Type.png)[](https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Linear_Regression1.png)Nell'equazione precedente, b 0 e b 1 sono i parametri del modello attuale che definiscono l'esatta sensibilità della variabile dipendente ai fattori esplicativi, e ε è la quantità di variabilità che non è spiegata dal modello. In pratica, tuttavia, questi valori esatti non sono noti con certezza e devono essere stimati dai dati.

L'equazione di regressione semplice corrispondente utilizzata per stimare i parametri è

Y = b 0+ b 1X dove:

Expected Value: E [b 0] = b 0\* E [b 1] = b 1 \* E [ε] = 0

Varianza: Var [ε] = σε2

Nel caso in cui abbiamo più fattori esplicativi, l'analisi è denominata modello di regressione multipla e ha la forma:

Y = b 0+ b 1X+ b 2X2 + ⋯ b kXk + ε

L'obiettivo dell'analisi di regressione è determinare l'insieme di fattori esplicativi e sensibilità corrispondenti che spieghino il più possibile i valori dipendenti osservati. Metriche di valutazione del modello Nell'effettuare l'analisi di regressione, le metriche importanti per analizzare l'analisi sono:

* b k = Parametro del modello - sensibilità stimata di y al fattore k
* SE (b k) = errore standard del parametro stimato b k
* σy = errore standard del modello di regressione
* R2 = bontà di adattamento (la percentuale della varianza complessiva spiegata dal modello)
* T-stat = valore critico per il parametro stimato
* F-stat = valore critico per l'intero modello

Presupposti del modello

Esistono sei ipotesi o proprietà principali di un modello di regressione lineare. Se qualcuno di questi presupposti viene violato, i risultati dell'analisi potrebbero essere sospetti e potenzialmente fornire informazioni errate sulla reale relazione tra una variabile dipendente e i suoi fattori. In questi casi, gli analisti devono apportare modifiche ai dati. I principali presupposti del modello di regressione lineare sono:

* Relazione lineare tra variabile dipendente e fattori esplicativi.
* Termine di errore medio zero: il valore atteso del termine di errore è zero. E (ε) = 0.
* Variazione costante: ogni termine di errore ha la stessa varianza, ad esempio, nessuna eteroschedasticità. Var (εk) = σ2 per tutti i k.
* Termini di errore non correlati, ad es. nessuna autocorrelazione. E (εk\*εk - t) = 0 per tutti i periodi di tempo ritardati t.
* Gli errori sono indipendenti da fattori esplicativi. Cov (ε, Xk) = 0 per tutti i fattori k.
* I fattori esplicativi sono indipendenti. Cov (Xj, Xk) = 0 per tutti i fattori j e k.