

Altri concetti sulla regressione

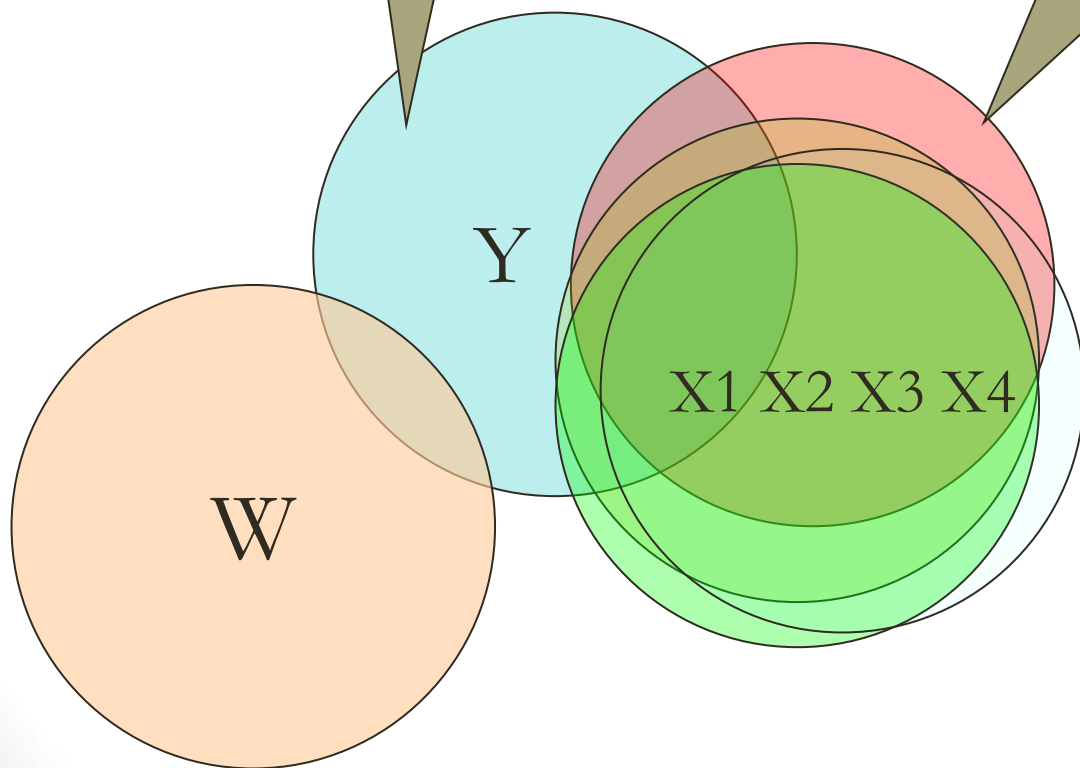
- Giovanni Battista Flebus
 - Lezioni di psicometria

1) Multicollinearità

- Varianza comune fra le VI: se è molto elevata produce stime instabili.
- Ci sono degli indici per indicare quando si manifesta questo effetto (tolleranza e VIF, Variance Inflation Factor)
- Soluzione: trasformare le variabili molto correlate in Componenti principali

Y- dipendente

X1-X2-X3-X4
Indipendenti molto
correlate



2) Importanza dei singoli casi

Nel campione esaminato, ci possono essere dei valori osservati che sono molto diversi dagli altri (Outliers, o **valori anomali**)

Per apprezzare l'apporto di ogni singola osservazione, si esaminano degli indici

1. Distanza di Mahalanobis
2. Distanza di Cook
3. Leverage

2.1 Distanza di Mahalanobis

- Una misura di distanza del punto k dagli altri punti, sulle variabili indipendenti

$$D_M(x) = \sqrt{(x - \mu)^T S^{-1} (x - \mu)}.$$

- **è un valore sempre positivo che indica la distanza di un punto di k dimensioni dalla media delle K dimensioni. Tiene conto sia della distanza di ogni k -esima media sia della varianza e della covarianza delle k variabili. Valori elevati indicano una grande distanza del punto da tutte le medie delle k variabili**

2.2 Distanza di Cook

- Quantificazione dell'effetto che avrebbe l'eliminazione del punto k sul calcolo dei residui. Valori elevati indicano che il punto k è un valore anomalo e richiede esame approfondito

2.3 Valore di influenza (leverage)

- È un altro indice, che dipende dalla distanza di Mahalanobis