

Q&A Webinar dell'11 marzo 2021

**Roberto G.**

Non ho capito esattamente come sono delineati i quartili nel box plot.

*Il boxplot è così definito: il limite inferiore della scatola rappresenta il primo quartile (0.25), la linea spessa che è dentro al boxplot è la mediana (0.5) ed il limite superiore il terzo quartile (0.75). Fuori dai "baffi" si trovano i valori anomali (circoletti) ed estremi (stelline).*

Il valore di PCA inferiore all'80 dell'esempio e' riferito alla variabile 1 o all'utilizzo di 1 sola variabile?

*Il valore dell'esempio visto inferiore all'80% è riferito alla variabilità spiegata dalla sola prima componente principale.*

**Lucia S.**

Fino a che valore cumulativo posso accettare che l'analisi PCA spieghi la correlazione dei miei dati?

*Il valore di varianza spiegata dalle prime "x" componenti principali deve essere almeno il 75%-80%.*

**Lorena S.**

Si può utilizzare la PCA avendo nel dataset 8 variabili? qual è il limite di numero di variabili?

*Non c'è limite al numero di variabili. Deve però sempre valere che il numero di osservazioni deve essere maggiore al numero delle variabili.*

**Paolo D.**

Quale criterio uso per scegliere le variabili da eliminare se voglio ridurre il loro numero?

L'autovalore lo uso sia per le componenti sia per le variabili?

*I criteri più utilizzati sono quelli dell'autovalore e della varianza spiegata. L'autovalore serve esclusivamente per le componenti principali.*

**Cristina P.**

Come mi comporto quando le unità di misura delle variabili sono molto diverse tra loro?

*Si possono standardizzare i dati in fase preliminare.*

**Simone Domenico G.**

Come faccio a stabilire quali variabili sarebbe meglio eliminare dall'analisi per ottenere una migliore percentuale cumulativa di spiegazione della diversità?

*Per guardare quali variabili sono "trainanti" in ogni componente principale basta guardare i loading che sono >0.7 nella matrice dei "ruotata" (ad es. con varimax).*

**Nicolaja G.**

Ci può fare altri esempi di dataset in cui si utilizza la PCA? per esempio in ambito clinico

*Quando si ha a che fare con molte variabili quantitative magari relativamente ad esami clinici o agli esiti di studi randomizzati.*

**Martina C.**

Gli ordini di grandezza/unità di misura delle variabili originali influenzano le CP? Se sì, come procediamo?

*Si standardizza in fase preliminare.*

**Marco V.** Se ho capito bene il concetto, la PCA è comunque applicabile negli studi clinici e come? Mi riferisco per esempio a studi clinici dove ho più frequentemente tante variabili con poche osservazioni (es. 100 pazienti reclutati -osservazioni- e 50 items per paziente quindi 5000 variabili)  
*In questo caso NO, si hanno troppe variabili.*

**Maria Cristina C.**

È possibile avere un plot tridimensionale con es CP1 e CP3 o CP2 e CP3?  
*Sì, Statistics lo permette.*

**Laura M.**

Può essere utilizzata con misure ripetute?  
*Meglio di no; l'ideale è prendere sempre dei campioni a cui applicare la PCA in quei casi.*