SPSS

Gold Business Partner

IBM

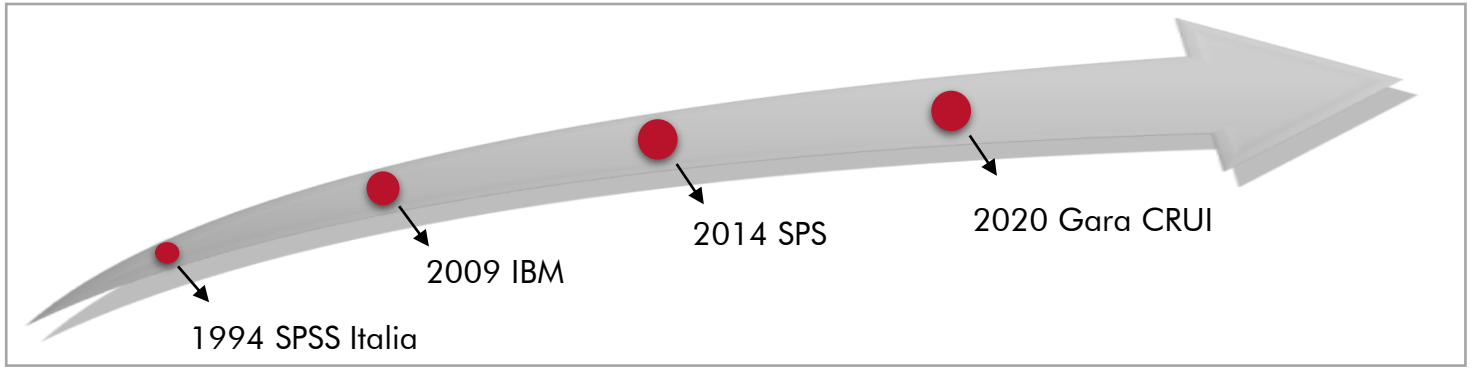
Competency
Data Science & Business Analytics

Webinar

L'Analisi delle Componenti Principali (PCA) con IBM SPSS Statistics

Webinar | SPS Srl

Roadmap



Posizionamento

- +25 anni di esperienza sui prodotti SPSS
- IBM BP **Software Support Provider** con competenza IBM Expert in Data Science and Business Analytics
- Gold Business Partner IBM

Presenza sul mercato

- **Scuole e Università**
- **Enti di ricerca (CNR, ISS, CREA ecc.)**
- Sanità (IRCCS, AO, AOU ecc.)
- Pubblica Amministrazione
- Associazioni, Fondazioni
- Aziende (GdO, Finance, Retail, ecc.)

Webinar | CRUI e SPS S.r.l.

Gold Business Partner

Il contratto stipulato tra CRUI e SPS S.r.l. in data 21 ottobre 2019 si riferisce alla fornitura licenze Campus IBM SPSS Statistics del "Catalogo licenze CRUI Campus PA" e di licenze aggiuntive IBM SPSS Amos e IBM SPSS Statistics per usi amministrativi.

Tutte le informazioni sul contratto sono disponibili al link

https://www.spss.it/crui-universita-italiane

Il referente commerciale per l'esecuzione contratto in SPS Srl è:

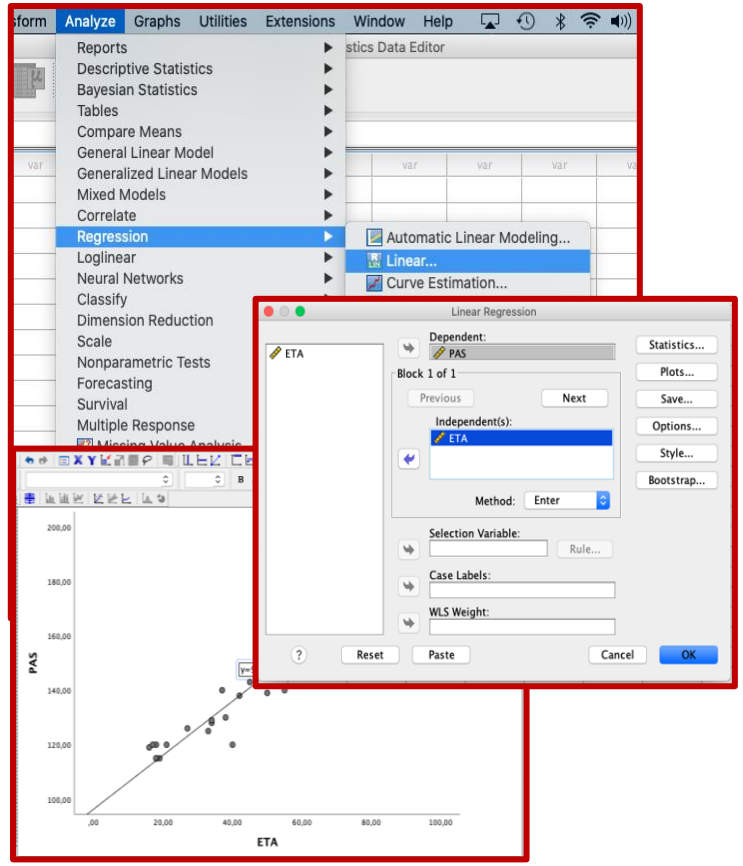
Laura Zerbini (+39 335 1360538, laura.zerbini@spss.it)

Laura Zerbini

Consulente in Italia di soluzioni SPSS per l'analisi statistica de...

Webinar | IBM SPSS Statistics al lavoro...

Gold Business Partner



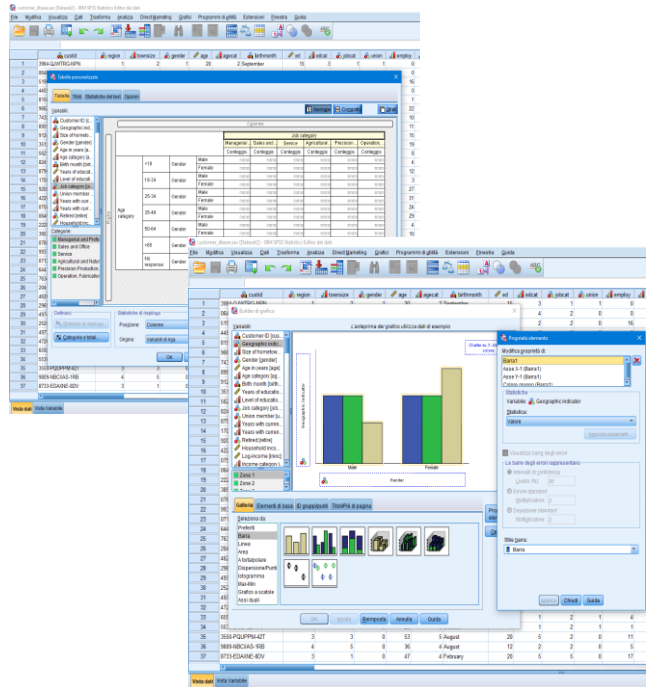
Mattia Cefis docente @SPS

www.spss.it

Webinar | Agenda

Gold Business Partner

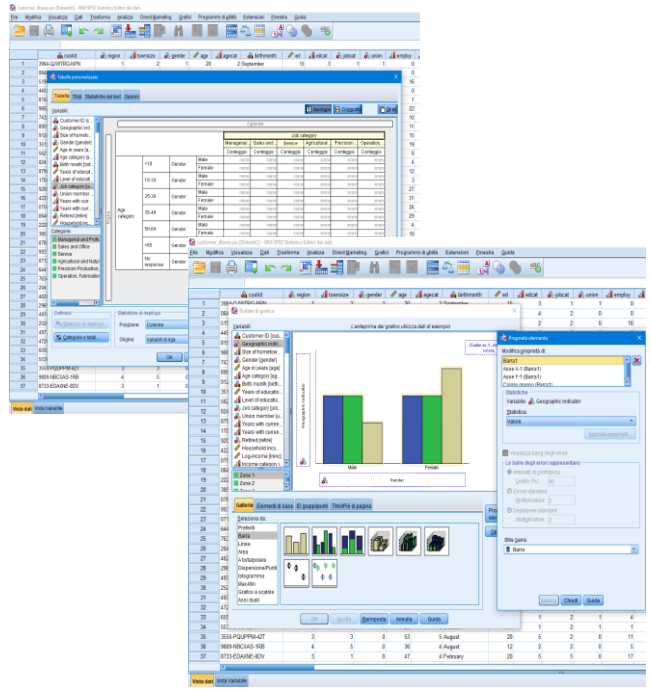
- Contesto introduttivo
- Teoria PCA
- Esempio con IBM SPSS Statistics
- Ringraziamenti



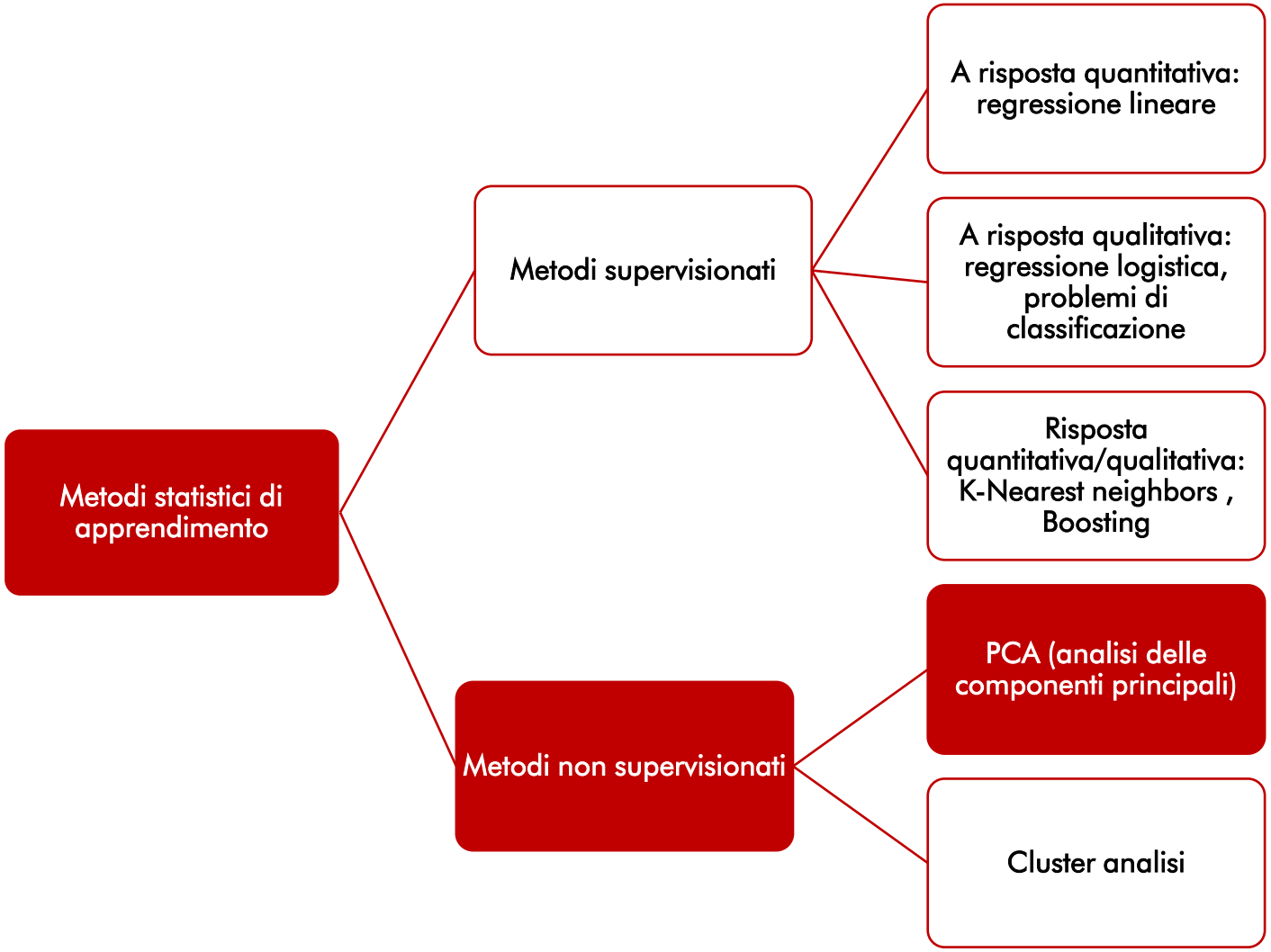
Webinar | Agenda

Gold Business Partner

- Contesto introduttivo
- Teoria PCA
- Esempio con IBM SPSS Statistics
- Ringraziamenti



Contesto introduttivo



Contesto introduttivo

PCA «goal»

- Ridurre il numero di variabili
- Mantenere una variabilità «spiegata» elevata

Contesto introduttivo

Gold Business Partner

Quando la PCA?

1)Presenza di correlazione tra le variabili

2)n°oss.>>n°var.

Contesto introduttivo

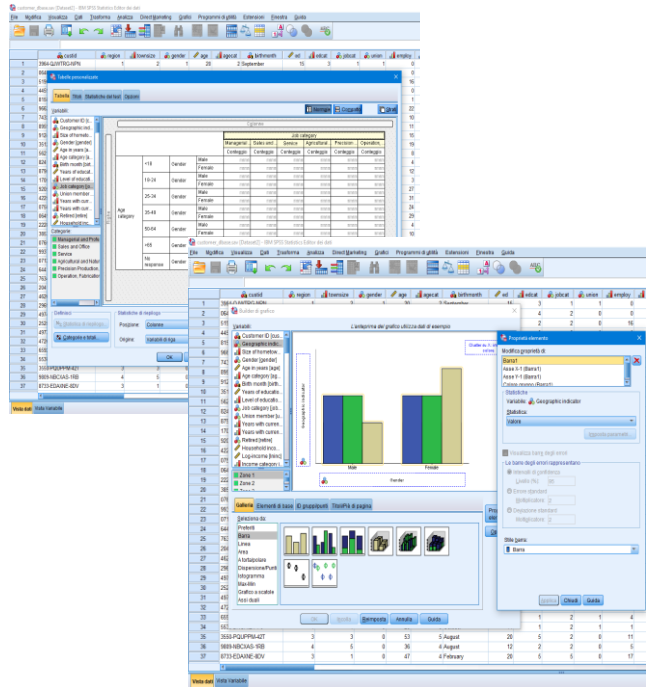
La PCA è una tecnica che ha lo scopo di ridurre il numero di variabili quando queste mostrano un pattern di correlazione tra loro. Ha molte applicazioni, ad esempio:

- In ambito **psicometrico**: si è interessati a ridurre il numero di item relativi ad un questionario.
- In ambito **scientifico**: quando si ha a che fare con dataset di grandi dimensioni e si evidenzia un certo pattern di collinearità tra le variabili presenti nell'analisi.
- In ambito **marketing**: analisi preliminare per identificare i cosiddetti pattern per variabili latenti di gradimento relativi ad un prodotto (e.g. sapore, profumo, percezione del marchio, ….)
- In ambito **sportivo**: analisi preliminare per capire quali pattern di variabili «guidano» l'analisi della performance di un atleta (e.g. forza, resistenza, skill mentali,…).

Webinar | Agenda

Gold Business Partner

- Contesto introduttivo
- Teoria PCA
- Esempio con IBM SPSS Statistics
- Ringraziamenti

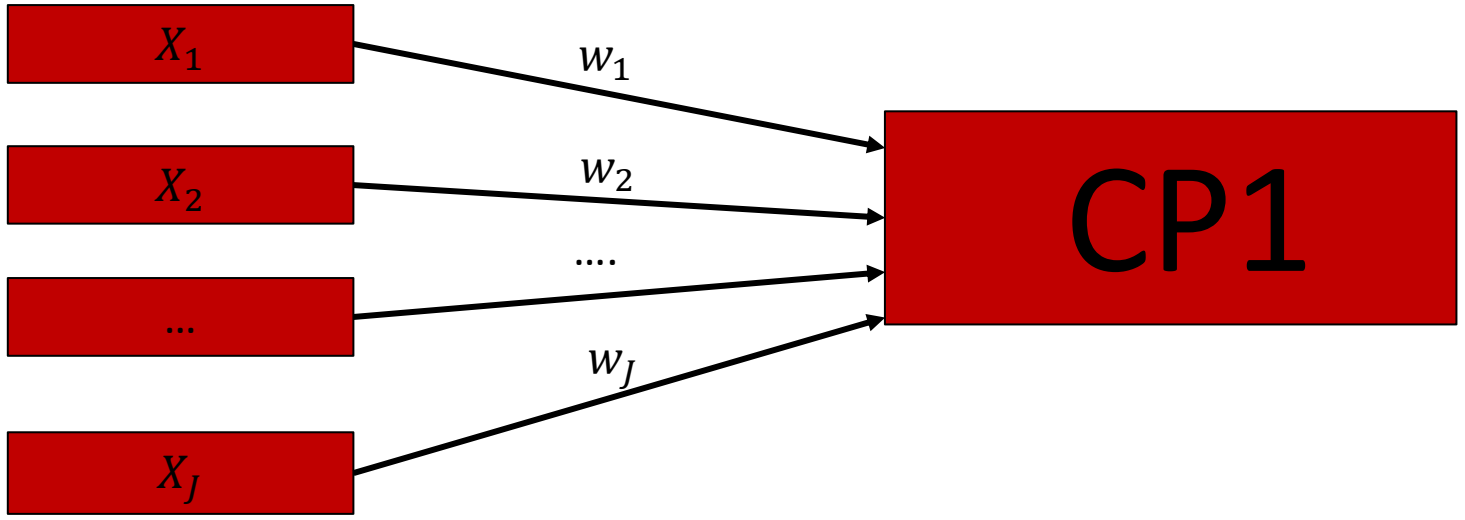


Teoria PCA

Obiettivo matematico/statistico: identificare quelle combinazioni lineari delle variabili iniziali che permettono di ridurre la dimensione del mio dataset iniziale e al tempo stesso mantenere una quota medio-alta di variabilità «spiegata» (quindi perdere meno informazione possibile).

Teoria PCA

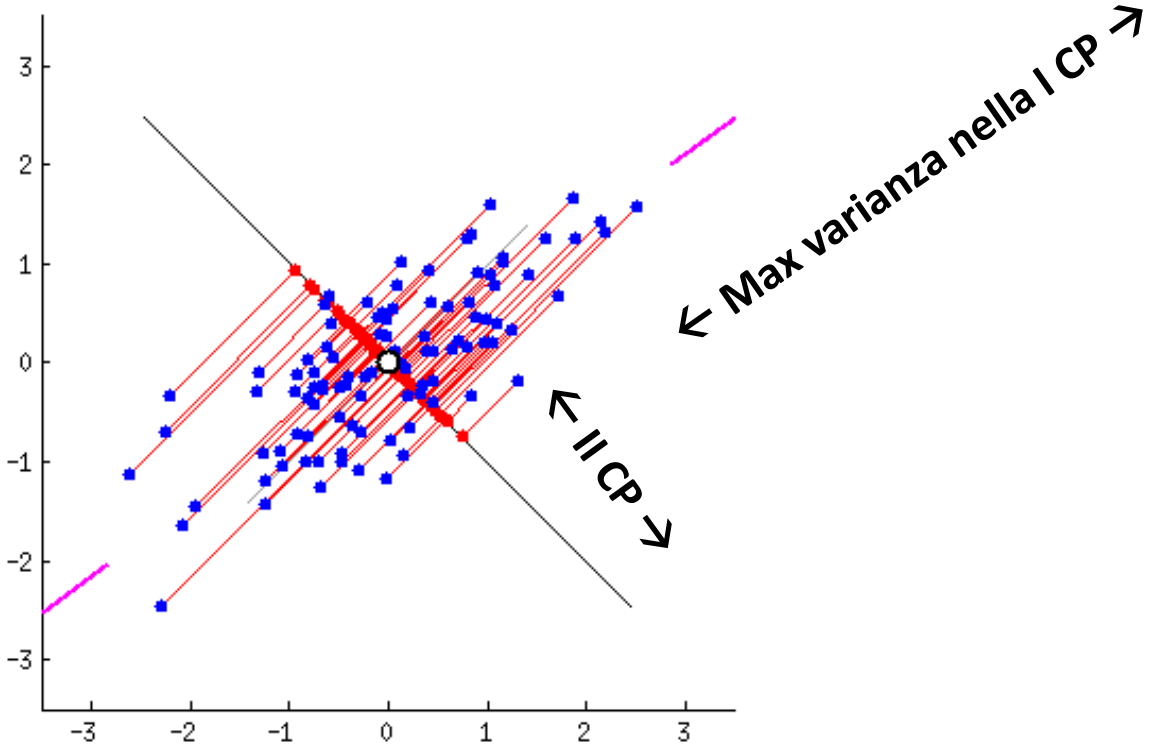
NB: le CP sono incorrelate tra loro! (e.g. Cov(CP1, CP2)=0)



$$CP1 = w_1 * X_1 + w_2 * X_2 + \ldots + w_J * X_J$$

Teoria PCA

Obiettivo geometrico: individuare uno spazio di dimensione ridotta su cui proiettare la nube dei punti originali in modo tale da deformare il meno possibile le distanze tra i punti.



Teoria PCA

1)I dati sono adatti per applicare la PCA?

- Se n°osservazioni >> n°variabili
- Se KMO test >0.5
- Se il test di sfericità di Bartlett è significativo (e.g. p-value<0.05)

Teoria PCA

2)Criteri per scegliere il n° adatto di CP:

- Se autovalore > 1
- Variabilità cumulata > 80%
- «Angolo» dello scree-plot

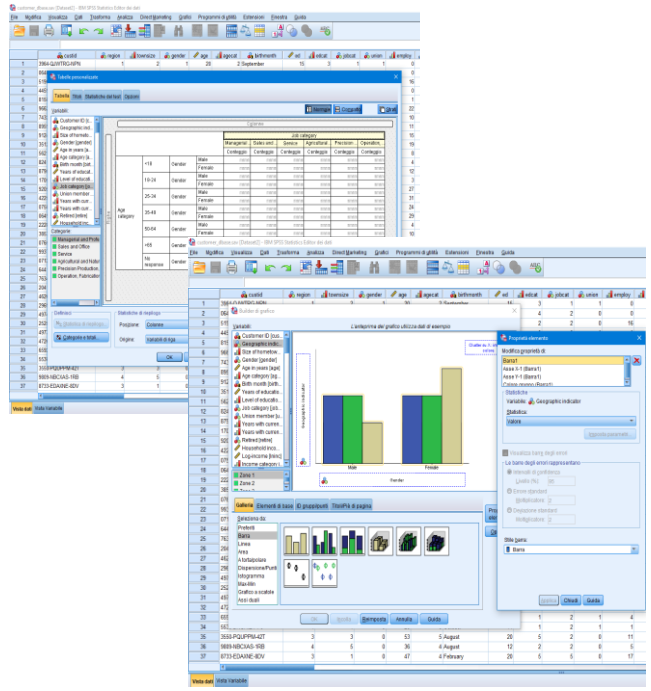
Teoria PCA

Angolo dello scree-plot: sceglierò 2 componenti principali!

Webinar | Agenda

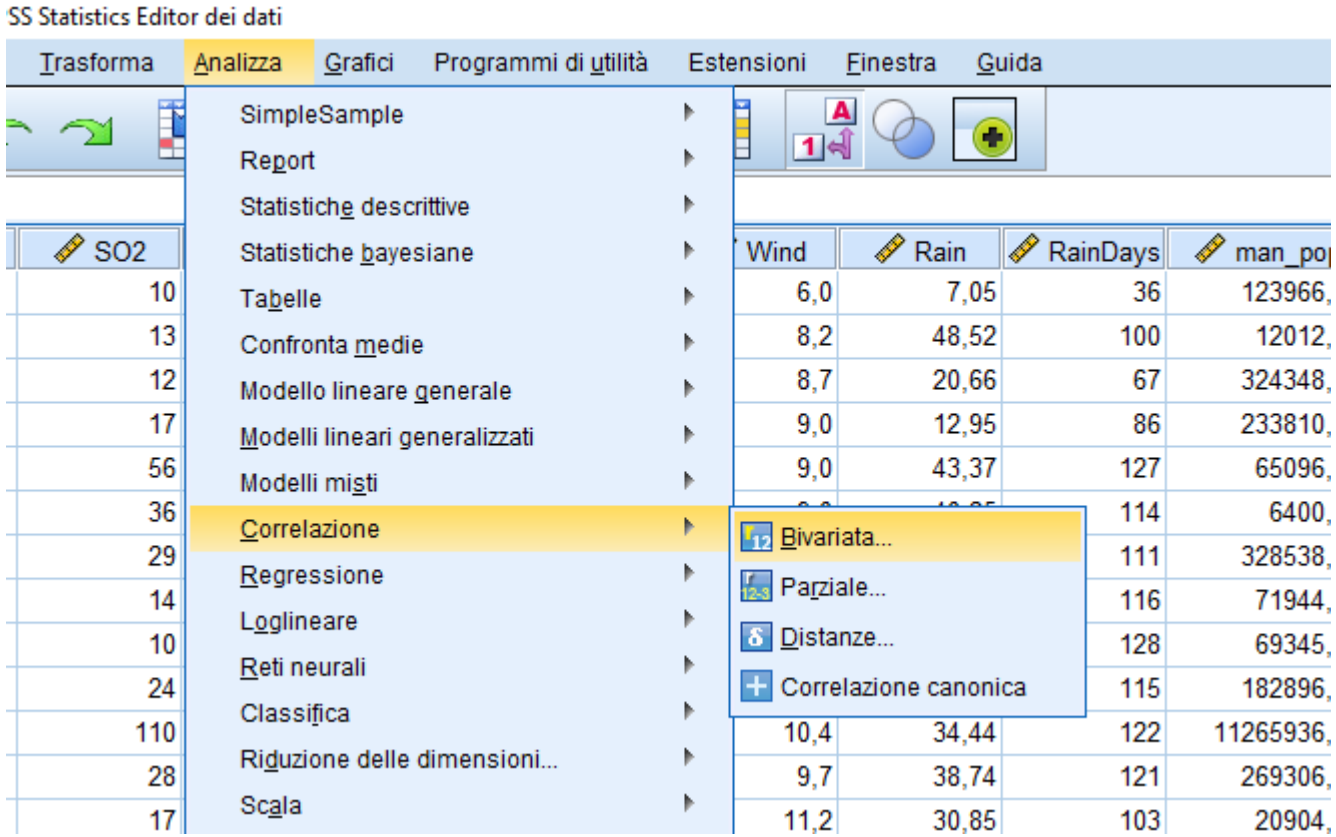
Gold Business Partner

- Contesto introduttivo
- Teoria PCA
- Esempio con IBM SPSS Statistics
- Ringraziamenti



Applicazione pratica

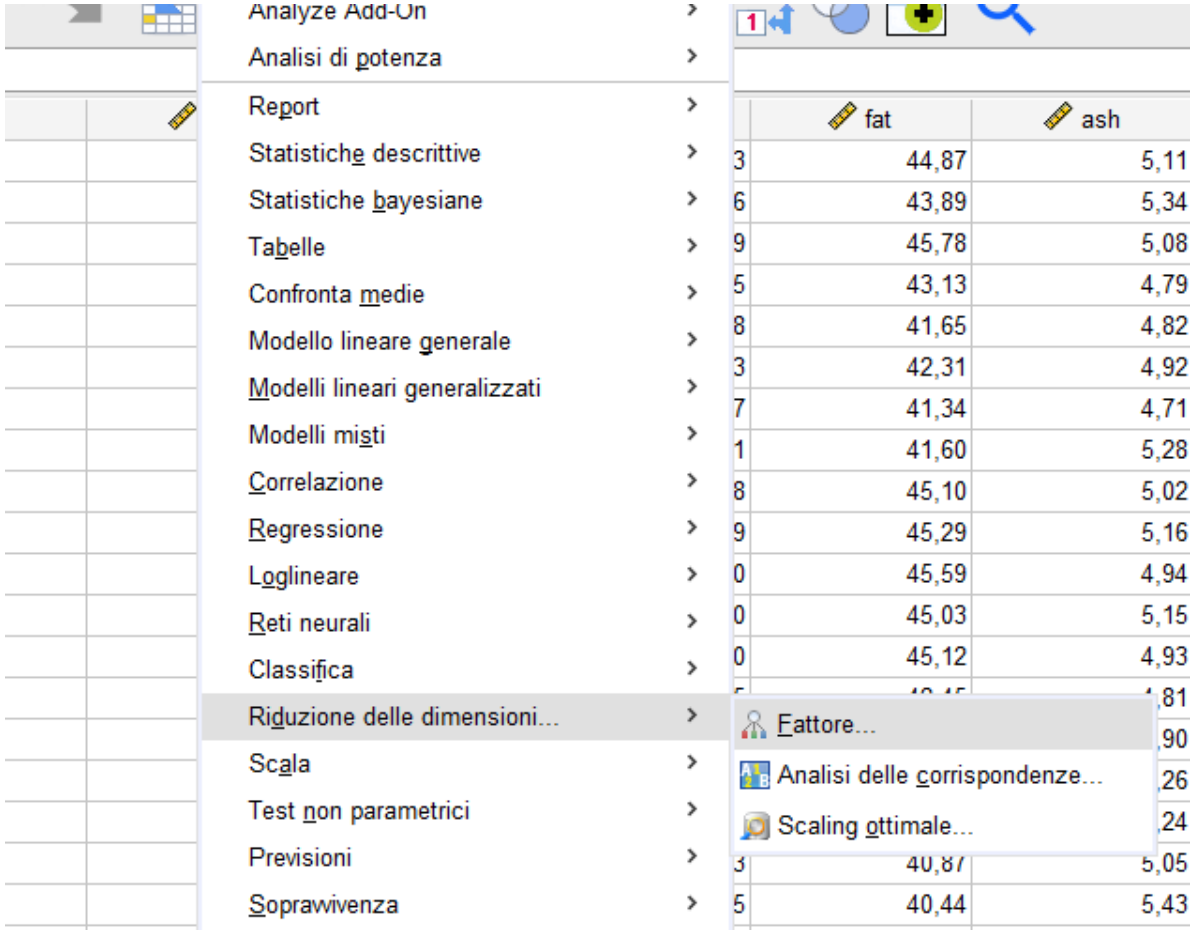
1)Analizza→ correlazione→bivariata



Applicazione pratica

Gold Business Partner

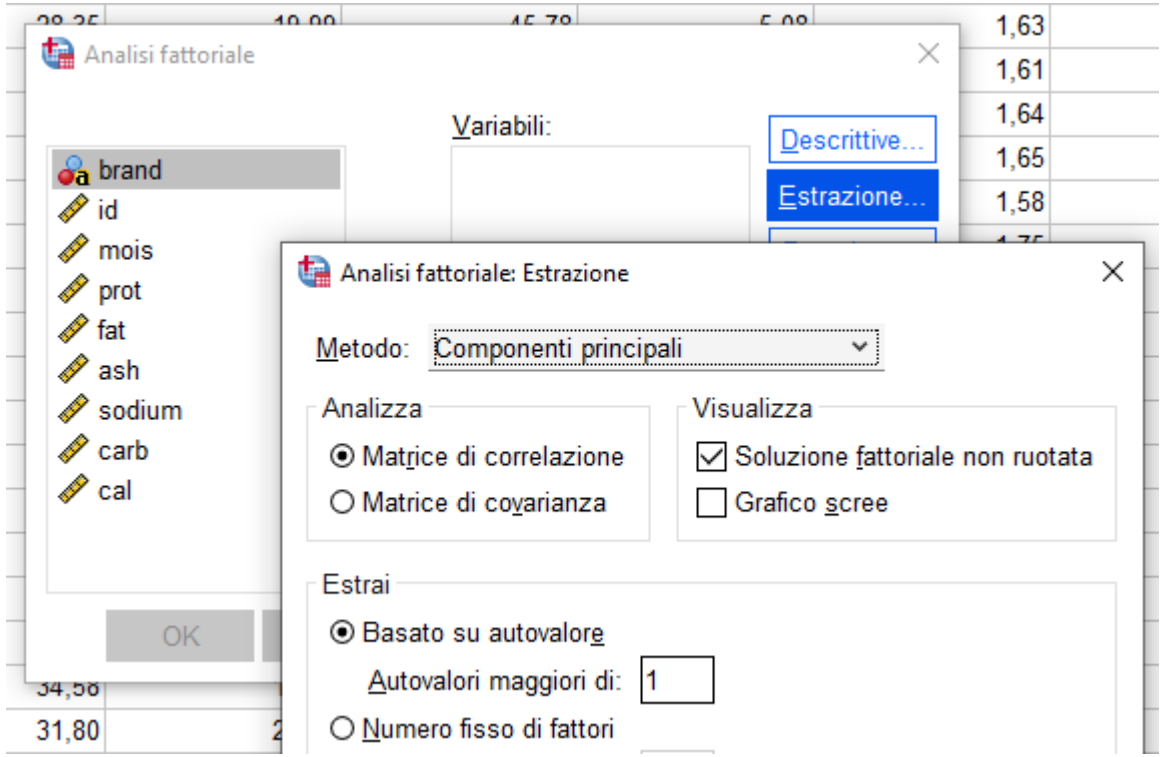
2)Analizza→ Riduzione dimensioni→ Fattore



www.spss.it

Applicazione pratica

2.1)Menù:



Applicazione pratica

3.1)Esempio output: test di verifica

Test di KMO e Bartlett

Misura di Kaiser-Meyer-Olkin di adeguatezza del campionamento.		,455
Test della sfericità di Bartlett	Appross. Chi-quadrato	7135,057
	gl	21
	Sign.	,000

Applicazione pratica

3.2)Esempio di output: autovalori e scelta del numero adatto di componenti principali

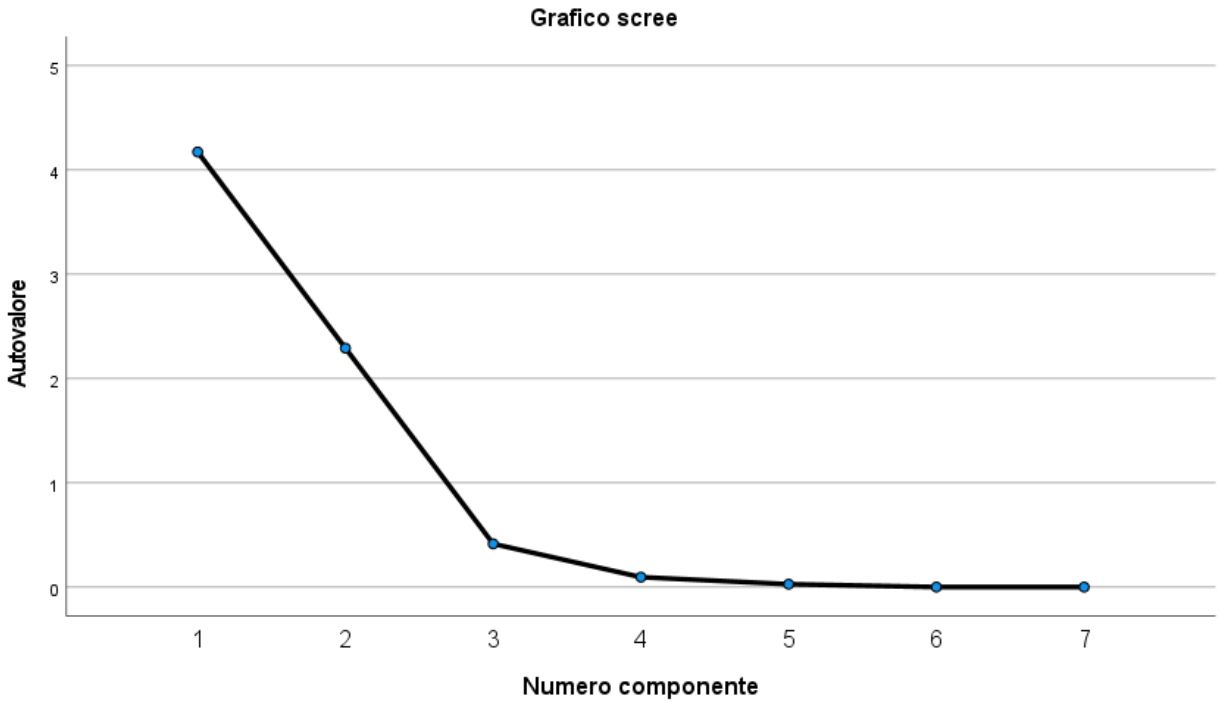
Sceglierò 2 componenti!

	Autovalori iniziali		
Componente	Totale	% di varianza	% cumulativa
1	4,172	59,597	59,597
2	2,290	32,721	92,318
3	,415	5,922	98,240
4	,095	1,360	99,600
5	,028	,395	99,995
6	,000	,005	100,000
7	9,519E-6	,000	100,000

Metodo di estrazione: Analisi dei componenti principali.

Applicazione pratica

3.3)Esempio di output: scree-plot



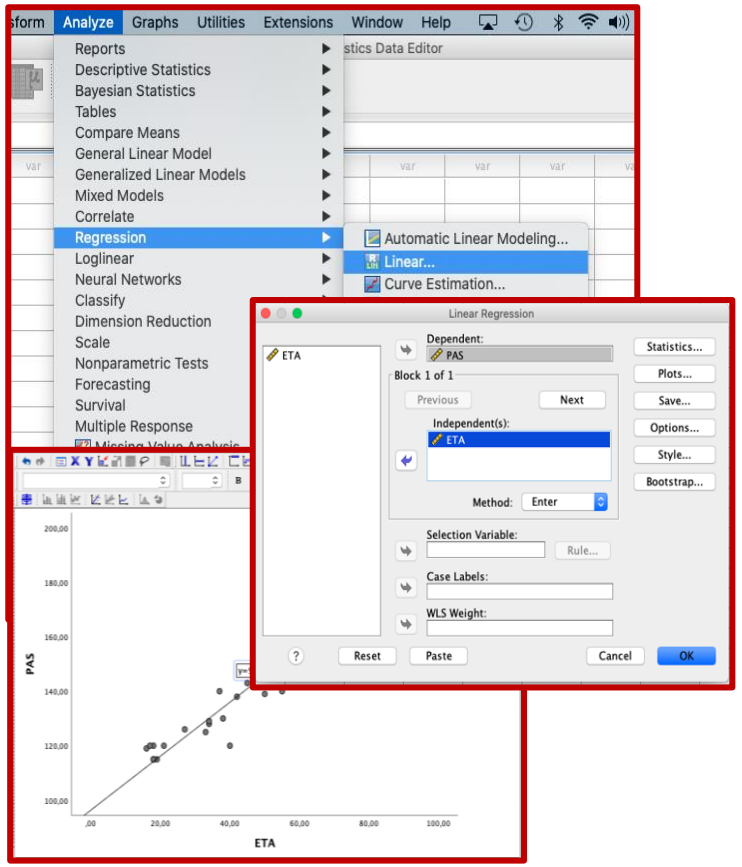
Conclusione

- ✓ OK con variabili quantitative
- ✓ OK quando sono presenti correlazioni rilevanti
- ✓ OK autovalori > 1
- ✓ Riduzione dimensionale
- ✓ Punto di partenza per ulteriori analisi

Webinar | PCA

Gold Business Partner

@Laura Zerbini



RINGRAZIAMENTI

www.spss.it

Grazie per aver partecipato

Gold Business Partner

Il referente commerciale in SPS Srl per scuole e università e per il contratto Campus CRUI è Laura Zerbini (+39 335 1360538, laura.zerbini@spss.it)

Non esitate a contattarmi per qualunque necessita!

Laura Zerbini

Consulente in Italia di soluzioni SPSS per l'analisi statistica de...

www.spss.it