

Vediamo ora alcuni esempi.

### **Esempio1 (titoli)**

Consideriamo alcuni dati riguardanti il tasso settimanale di ritorno di cinque titoli della borsa di New York, registrati da Gennaio 1975 a Dicembre 1976.

Il tasso di ritorno un tasso di variazione del prezzo, da una settimana all'altra.

E' definito come:

$$\frac{(\text{prezzo di chiusura del Venerdì} - \text{prezzo di chiusura del Venerdì precedente})}{(\text{prezzo di chiusura del Venerdì precedente})}$$

I cinque titoli sono: Allied Chemical, du Pont, Union Carbide, Exxon e Texaco.

Disponiamo di sono 100 osservazioni, per ciascun titolo.

Si conduca:

- 1) Una analisi esplorativa dei dati (medie, varianze, ecc.) e si interpretino i risultati
- 2) Si riassume in maniera ragionevole l'informazione, e si rappresenti e interpreti l'andamento della borsa (ACP su spazio variabili)
- 3) Si interpreti l'informazione fornita dai dati (ACP su spazio unita').

Iniziamo con l'analisi esplorativa:

<b>Variable</b>	<b>N</b>	<b>Mean</b>	<b>StDev</b>
allied c	100	0.00543	0.04037
du pont	100	0.00483	0.03506
uni carb	100	0.00565	0.03945
exxon	100	0.00629	0.02833
texaco	100	0.00371	0.02755

La exxon sembra il titolo migliore, avendo media del tasso di ritorno piu' alta e deviazione standard bassa (al penultimo posto).

## Facciamo ora una analisi in componenti principali.

Dato che le variabili hanno la stessa unità di misura l'analisi potrà essere condotta sulla matrice S.

### Eigenvalues of the Covariance Matrix

	Autovalore	Differenza	Proporzione	Cumulata
1	0.00359539	0.00280321	0.6016	0.6016
2	0.00079218	0.00005574	0.1326	0.7341
3	0.00073644	0.00022777	0.1232	0.8574
4	0.00050867	0.00016490	0.0851	0.9425
5	0.00034377		0.0575	1.0000

Decidiamo di utilizzare solo le prime tre componenti principali, perché spiegano circa l'85% della varianza. Decidiamo inoltre di condurre l'analisi sulla matrice di varianze e covarianze.

Queste componenti sono:

Variable	PC1	PC2	PC3
allied c	-0.561	0.739	0.126
du pont	-0.470	-0.093	0.468
uni carb	-0.547	-0.654	0.114
exxon	-0.291	-0.113	-0.610
texaco	-0.284	0.071	-0.617

I coefficienti riportati ci permettono di interpretare ciascuna componente riportata (ACP su spazio unita')

La prima componente ha segni concordi. E' un indice di andamento generale della borsa. Notare che le compagnie petrolifere hanno molto meno peso delle altre tre.

La seconda componente contrasta la allied con la uni carb. Maggiori informazioni su questi titoli possono permettere interpretazioni interessanti. I due titoli a prima vista sono simili, ma potrebbe essere che quando uno e' in perdita l'altro guadagna punti percentuali. Maggiore analisi sarebbe in questo caso necessaria....

La terza componente contrasta i titoli petroliferi, piu' stabili; con i titoli meno stabili. E' piu' o meno una variabile che misura il rischio della borsa (se e' alto, conveniva investire a lungo termine su titoli stabili, se basso conveniva investire a breve termine su titoli instabili)

E se avessimo lavorato sulla matrice di correlazione?

I risultati sarebbero stati diversi anche se non troppo dissimili.

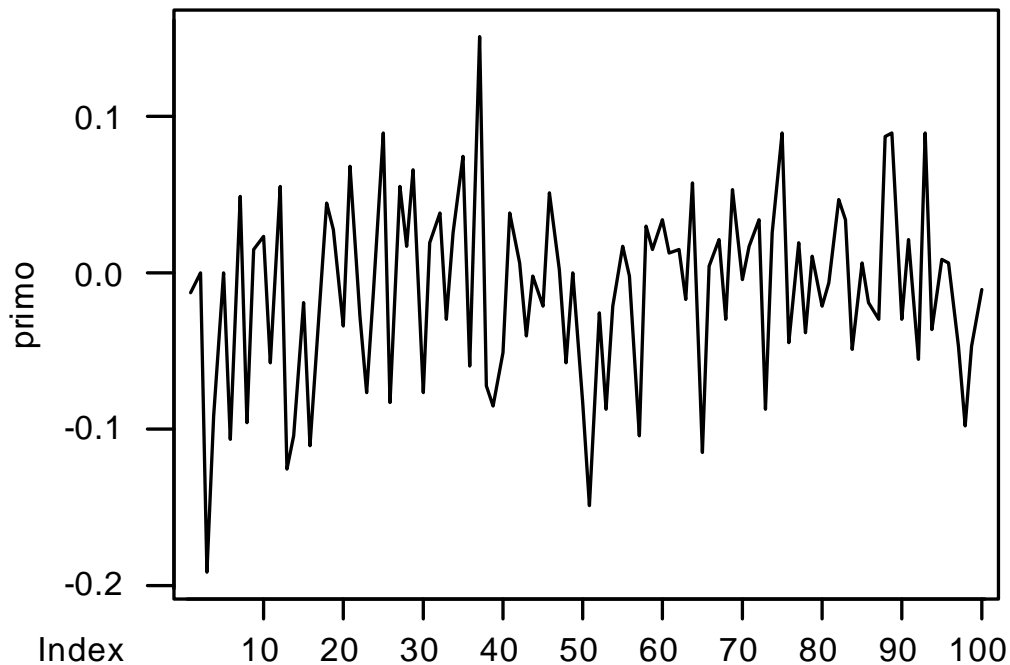
*Eigenvalues of the Correlation Matrix*

	<i>Autovalore</i>	<i>Differenza</i>	<i>Proporzione</i>	<i>Cumulata</i>
1	2.85648688	2.04736839	0.5713	0.5713
2	0.80911850	0.26907452	0.1618	0.7331
3	0.54004398	0.08869715	0.1080	0.8411
4	0.45134682	0.10834301	0.0903	0.9314
5	0.34300382		0.0686	1.0000

*Eigenvectors*

	<i>Prin1</i>	<i>Prin2</i>	<i>Prin3</i>
<i>alliedc</i>	- 0.463541	-.240850	-.613357
<i>_dupont</i>	- 0.457076	-.509100	0.177900
<i>_unicarb</i>	-0.469980	-.260577	0.337036
<i>_exxon</i>	-0.421677	0.525265	0.539018
<i>_texaco</i>	-0.421329	0.582242	-.433603

Rimane solo da utilizzare gli scores per avere un andamento medio della borsa. La prima componente principale spiega gran parte della varianza, indica piu' o meno un valore medio: e' un buon indice dell'andamento della borsa.



La borsa ha avuto un andamento sinusoidale, ma piu' o meno stabile per quelle 100 settimane.

Si noti che, poiche' il primo autovettore ha tutti coefficienti negativi, valori alti nella prima CP indicano un cattivo andamento.

Si noti un crollo di borsa (picco alto) intorno alla 35ma settimana e un momento di forte salita della borsa (picco basso) alla fine del primo mese circa.

In sostanza, la borsa ha avuto un andamento discendente nelle prime quaranta settimane, poi un momento di forti rialzi tra la quarantesima e la sessantesima, e poi e' stata piu' o meno stabile.

## Esempio2

I dati riguardano 48 persone candidate per un posto di lavoro in una multinazionale; i candidati vengono intervistati e classificati in base a 15 criteri:

CU: curriculum

APP: apparenza

AA: abilita' accademiche

GR: gradevolezza

FID: fiducia in se'

LUC: lucidita'

ON: onesta'

CDV: capacita' di vendere

ESP: esperienza

CC: capacita' di coinvolgere

AMB: ambizione

CI: capacita' di comprendere nuove idee

POT: potenziale di sviluppo

VOL: volonta' di essere assunti

AD: quanto e' adatto al posto

Ciascun criterio e' un punteggio da 0 a 10.

- 1) Quali misure/criteri sono importanti? (ACP su spazio osservazioni)
- 2) Chi assumiamo? (ACP su spazio variabili)
- 3) Quali misure/criteri sono importanti? (ACP su spazio osservazioni)
- 4) Chi assumiamo? (ACP su spazio variabili)

Dato che le unità di misura delle variabili coincidono (sono punteggi da 0 a 100) possiamo effettuare l'analisi sulla matrice di varianze e covarianze.

### Eigenvalues of the Covariance Matrix

	Autovalore	Differenza	Proporzione	Cumulata
1	66.5364216	48.3558875	0.5430	0.5430
2	18.1805340	7.5895485	0.1484	0.6914
3	10.5909855	3.8230376	0.0864	0.7778
4	6.7679478	2.7823032	0.0552	0.8330
5	3.9856446	0.3579742	0.0325	0.8656
6	3.6276704	0.7119224	0.0296	0.8952
7	2.9157480	0.0802174	0.0238	0.9190
8	2.8355306	0.8804081	0.0231	0.9421
9	1.9551225	0.3416313	0.0160	0.9581
10	1.6134912	0.4770192	0.0132	0.9712
11	1.1364720	0.2636018	0.0093	0.9805
12	0.8728702	0.1661951	0.0071	0.9876
13	0.7066751	0.1981398	0.0058	0.9934
14	0.5085353	0.2071663	0.0042	0.9975
15	0.3013690		0.0025	1.0000

Le prime quattro componenti spiegano quasi l'83% della variabilità. Dato che le componenti seguenti non aumentano drasticamente la percentuale di varianza spiegata, decidiamo di utilizzare solo le prime quattro. Si noti che si è così passati da 15 a 4 variabili, con una notevole sintesi dei dati.

I coefficienti delle prime 4 componenti principali sono:

Variable	PC1	PC2	PC3	PC4
CU	-0.149	0.371	0.200	-0.277
APP	-0.132	-0.029	0.042	0.134
AA	-0.030	0.102	-0.131	0.603
GR	-0.203	-0.093	0.620	0.126
FID	-0.231	-0.236	-0.189	-0.072
LUC	-0.337	-0.196	-0.125	0.053
ON	-0.120	-0.301	0.447	0.256
CDV	-0.379	-0.090	-0.282	-0.172
ESP	-0.164	0.636	0.025	0.166
CC	-0.316	0.012	-0.113	-0.135
AMB	-0.312	-0.122	-0.245	-0.147
CI	-0.339	-0.074	-0.050	0.206
POT	-0.357	-0.025	0.041	0.317
VOL	-0.226	-0.045	0.385	-0.460
AD	-0.274	0.471	0.017	-0.016

I coefficienti ci permettono di interpretare l'informazione fornita dalle componenti.

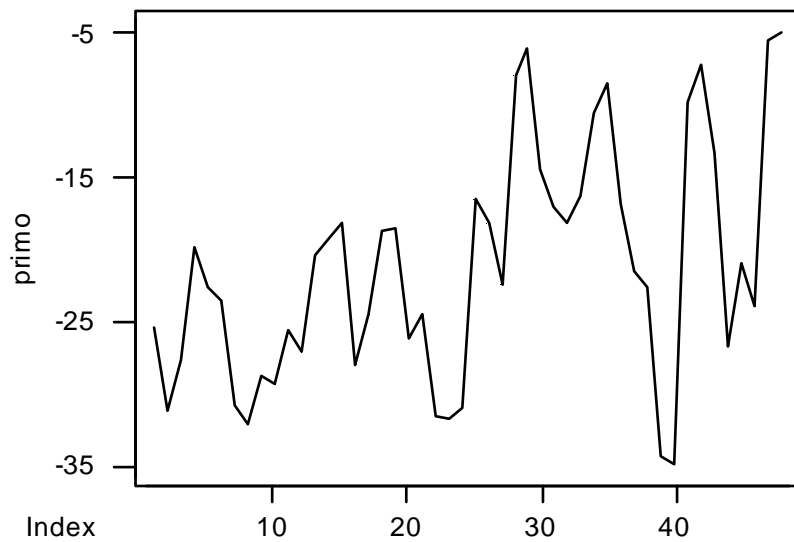
PC1, con coefficienti tutti dello stesso segno, e' una media ponderata dei punteggi nei criteri, con segno meno davanti.

Si noti che le abilita' accademiche (AA) non hanno praticamente peso.

Evidentemente, tutti i candidati piu' o meno si equivalgono. Differenze si osservano per la lucidita' (LUC), CDV, CC, AMB, CI, POT.

La seconda componente principale contrasta CU, ESP, AD con ON e FID. E' un indice interessante: alti valori indicano candidati poco onesti e/o sfiduciati ma con ottimo curriculum, esperienza e indicazione per il posto. Bassi valori indicano candidati molto onesti o con molta fiducia in se stessi ma poco adatti, inesperti o con scarso curriculum. Non e' un indice che si presta a interpretazioni dirette, ma magari sta a indicare che nei dati vi sono questo tipo di individui: i piu' adatti e/o esperti sono anche meno onesti e/o sfiduciati. Difficile dire se sia meglio assumere chi ha un valore alto o, viceversa, chi ha un valore basso!!

Ora dobbiamo decidere chi assumere. Ci sono vari modi per determinare chi ha un insieme di punteggi piu' alti. Si puo' fare un grafico della prima componente principale, o delle prime due.



si individuano chiaramente due individui con voti molto alti (osservazioni con PC1 piu' basso).

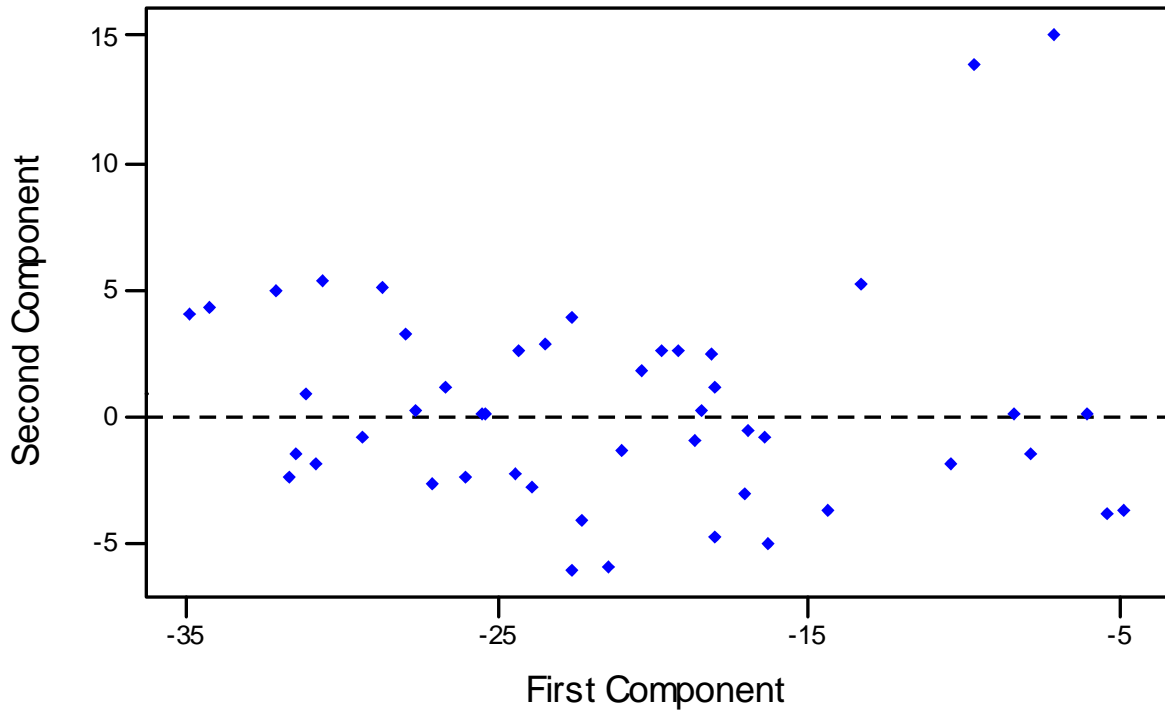
Pertanto gli individui intervistati al 39 e 40mo posto sono coloro meritevoli di assunzione.

NOTA:

E' possibile e ragionevole farsi i pesi da soli, a priori, e determinare un indice generale di assumibilita', in base al quale decidere chi assumere.



Score Plot of CU-AD



## Ultimo esempio

### Matrice dei dati: Paesi e consumi alimentari (dati espressi in kg pro-capite per anno)

	cereali	riso	patate	zucchero	verdure	vino	carne	latte	burro	uova
Belgio	72,20	4,20	98,80	40,40	103,20	20,90	102,00	80,00	7,70	14,20
Danimarca	70,50	2,20	57,00	39,50	50,00	22,00	105,80	145,20	4,10	14,30
Germania	71,30	2,30	74,10	37,10	83,10	22,80	97,20	90,70	6,90	14,80
Grecia	109,80	5,40	90,00	30,00	229,50	25,30	77,10	63,10	0,90	11,30
Spagna	71,40	5,80	107,80	26,80	191,70	43,00	102,10	98,40	0,60	15,30
Francia	73,00	4,30	78,20	34,10	95,00	64,50	110,50	98,90	8,90	15,00
Irlanda	93,40	3,20	151,50	34,80	55,00	3,90	105,00	185,90	3,40	11,40
Italia	110,20	4,80	38,60	27,90	181,90	61,60	88,00	65,00	2,40	11,10
Olanda	54,60	5,00	86,70	39,70	99,00	14,00	89,40	136,20	5,40	10,70
Portogallo	86,00	5,70	106,60	29,40	100,00	57,00	75,50	96,00	1,50	7,70
RegnoUnito	74,30	4,50	94,10	39,80	60,00	10,40	74,40	129,30	3,20	10,80
Austria	68,70	4,20	62,60	37,10	81,90	34,30	93,40	121,30	4,30	13,40
Finlandia	70,10	5,40	61,60	35,70	52,60	10,20	65,00	208,40	5,80	10,90
Islanda	79,70	1,90	50,20	54,90	50,00	6,20	71,70	205,60	4,60	11,30
Norvegia	76,90	3,50	73,20	37,30	48,30	6,60	54,90	176,50	2,10	11,30
Svezia	69,30	4,30	70,00	37,50	48,50	12,30	60,50	154,10	5,70	12,90

## Simple Statistics

	cereali	riso	patate	zucchero	verdure
Mean	78.21250000	4.168750000	81.31250000	36.37500000	95.60625000
StD	14.92603430	1.242963529	27.34817179	6.61316868	56.81797214

### Simple Statistics

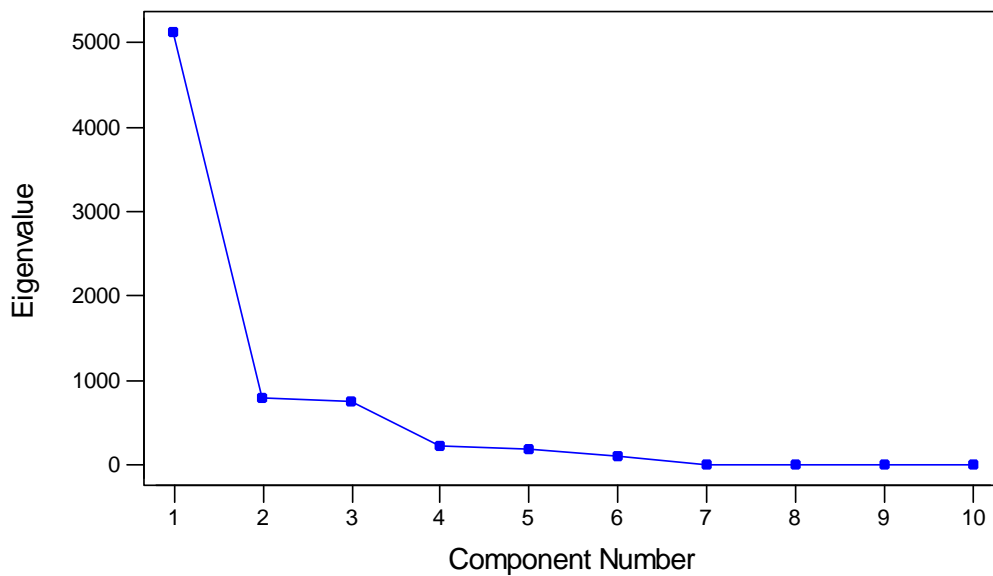
	vino	carne	latte	burro	uova
Mean	25.93750000	85.78125000	128.4125000	4.218750000	12.27500000
StD	20.33499037	17.60769600	47.5408579	2.428777127	2.07990384

**Si nota che alcune variabili hanno varianza molto piu' elevata di altre (verdure, latte, patate)**

## Eigenvalues of the Covariance Matrix

	<b>Autovalore</b>	<b>Differenza</b>	<b>Proporzione</b>	<b>Cumulata</b>
<b>1</b>	<b>5132.92829</b>	<b>4335.75101</b>	<b>0.7091</b>	<b>0.7091</b>
<b>2</b>	<b>797.17728</b>	<b>39.85686</b>	<b>0.1101</b>	<b>0.8193</b>
<b>3</b>	<b>757.32042</b>	<b>522.90368</b>	<b>0.1046</b>	<b>0.9239</b>
<b>4</b>	<b>234.41674</b>	<b>52.43614</b>	<b>0.0324</b>	<b>0.9563</b>
<b>5</b>	<b>181.98060</b>	<b>66.81981</b>	<b>0.0251</b>	<b>0.9814</b>
<b>6</b>	<b>115.16079</b>	<b>100.57350</b>	<b>0.0159</b>	<b>0.9973</b>
<b>7</b>	<b>14.58729</b>	<b>11.76225</b>	<b>0.0020</b>	<b>0.9994</b>
<b>8</b>	<b>2.82505</b>	<b>1.21748</b>	<b>0.0004</b>	<b>0.9998</b>
<b>9</b>	<b>1.60756</b>	<b>1.44077</b>	<b>0.0002</b>	<b>1.0000</b>
<b>10</b>	<b>0.16680</b>		<b>0.0000</b>	<b>1.0000</b>

Scree Plot of cereali-uova



## Eigenvectors

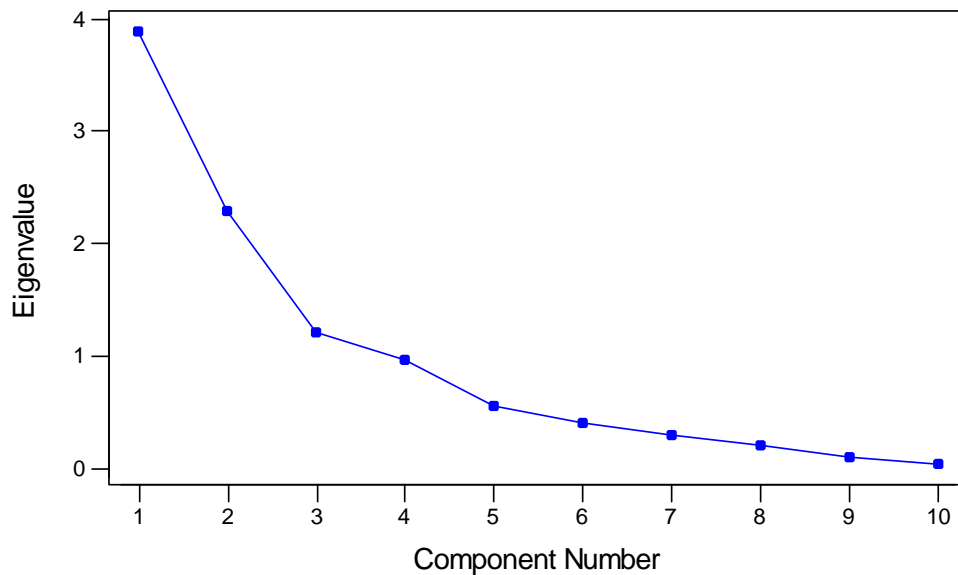
	Prin1	Prin2
cereali	0.106659	0.066292
riso	0.009112	0.009059
patate	0.026862	<b>0.906114</b>
zucchero	-.062354	-.047323
verdure	<b>0.757130</b>	0.205854
vino	0.191421	-.227049
carne	0.084218	0.105490
latte	<b>-.605649</b>	0.257544
burro	-.011138	-.027063
uova	0.003983	-.011152

**L'analisi effettuata sulla base della matrice delle covarianze non va bene. E' piu' opportuno lavorare sulla matrice di correlazione.**

## Eigenvalues of the Correlation Matrix

	<b>Autovalore</b>	<b>Differenza</b>	<b>Proporzione</b>	<b>Cumulata</b>
<b>1</b>	<b>3.89203172</b>	<b>1.59541317</b>	<b>0.3892</b>	<b>0.3892</b>
<b>2</b>	<b>2.29661854</b>	<b>1.08404624</b>	<b>0.2297</b>	<b>0.6189</b>
<b>3</b>	<b>1.21257231</b>	<b>0.24535378</b>	<b>0.1213</b>	<b>0.7401</b>
<b>4</b>	<b>0.96721852</b>	<b>0.40776650</b>	<b>0.0967</b>	<b>0.8368</b>
<b>5</b>	<b>0.55945203</b>	<b>0.15302642</b>	<b>0.0559</b>	<b>0.8928</b>
<b>6</b>	<b>0.40642560</b>	<b>0.10137598</b>	<b>0.0406</b>	<b>0.9334</b>
<b>7</b>	<b>0.30504962</b>	<b>0.09028603</b>	<b>0.0305</b>	<b>0.9639</b>
<b>8</b>	<b>0.21476359</b>	<b>0.10562019</b>	<b>0.0215</b>	<b>0.9854</b>
<b>9</b>	<b>0.10914340</b>	<b>0.07241872</b>	<b>0.0109</b>	<b>0.9963</b>

Scree Plot of cereali-uova



### Eigenvectors

	Prin1	Prin2	Prin3	Prin4
cereali	0.298783	-.236262	-.175283	<b>0.610674</b>
riso	<b>0.350267</b>	-.179432	0.127899	<b>-.619730</b>
patate	0.107756	-.046411	<b>0.858696</b>	0.099971
zucchero	<b>-.441430</b>	0.007159	-.146741	0.178337
verdure	<b>0.449846</b>	0.031505	-.107289	0.130004
vino	0.370344	0.233749	-.269239	-.174596
carne	0.122585	<b>0.536427</b>	0.289564	0.256022
latte	<b>-.401962</b>	-.257325	0.141491	-.022487
burro	-.253037	<b>0.422509</b>	-.078305	-.283054
uova	-.037769	<b>0.567155</b>	0.005145	0.087504

Prima componente principale e' legate principalmente a:  
zucchero, latte (con segno negativo)  
riso e verdure (segno positivo)

Il secondo autovettore ha coefficienti alti (e positivi) per carne  
burro e uova;  
la seconda CP e' legata alla quantità di proteine presenti nella  
dieta.

La terza CP e' legata soprattutto al consumo di patate .

La quarta a riso e cereali (in contrasto tra loro).

