

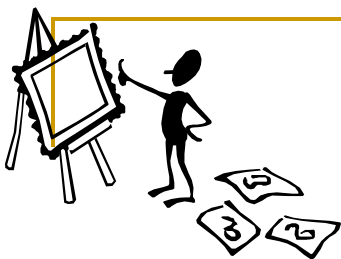
Analisi dei Dati 2



Esempio ACP: i consumi alimentari

16 paesi, 10 variabili continue

	Cere	Riso	Pata	Zucc	Verd	Vino	Carn	Latt	Burr	Uova
Belgio	72,20	4,20	98,80	40,40	103,2	20,90	102,0	80,00	7,70	14,20
Danimarca	70,50	2,20	57,00	39,50	50,00	22,00	105,8	145,2	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	109,8	5,40	90,00	30,00	229,5	25,30	77,10	63,10	0,90	11,30
Spagna	71,40	5,80	107,8	26,80	191,7	43,00	102,1	98,40	0,60	15,30
Francia	73,00	4,30	78,20	34,10	95,00	64,50	110,5	98,90	8,90	15,00
Irlanda	93,40	3,20	151,5	34,80	55,00	3,90	105,0	185,9	3,40	11,40
Italia	110,2	4,80	38,60	27,90	181,9	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	136,2	5,40	10,70
Portogallo	86,00	5,70	106,6	29,40	100,0	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	129,3	3,20	10,80
Austria	68,70	4,20	62,60	37,10	81,90	34,30	93,40	121,3	4,30	13,40
Finlandia	70,10	5,40	61,60	35,70	52,60	10,20	65,00	208,4	5,80	10,90
Islanda	79,70	1,90	50,20	54,90	50,00	6,20	71,70	205,6	4,60	11,30
Norvegia	76,90	3,50	73,20	37,30	48,30	6,60	54,90	176,5	2,10	11,30
Svezia	69,30	4,30	70,00	37,50	48,50	12,30	60,50	154,1	5,70	12,90



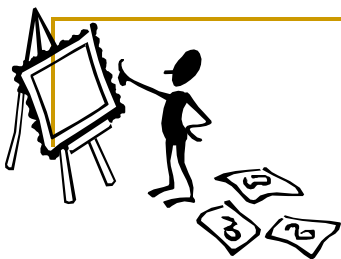
Analisi in R^p

(Spazio delle variabili)

$$x_{ij} = \frac{y_{ij} - \bar{y}_j}{s_j \sqrt{n}} \longrightarrow \mathbf{R} = \mathbf{X}'\mathbf{X} \text{ matrice di correlazione}$$
$$s_j = \sqrt{(1/n) \sum_i (y_{ij} - \bar{y}_j)^2}$$

Matrice di correlazione:

	cereali	riso	patate	zucch.	verd.	carne	latte	burro	uova
cereali	1,00								
riso	0,13	1,00							
patate	0,06	0,23	1,00						
zucchero	-0,41	-0,69	-0,28	1,00					
verdure	0,56	0,57	0,07	-0,64	1,00				
carne	-0,07	-0,15	0,29	-0,19	0,22	1,00			
latte	-0,34	-0,39	-0,04	0,58	-0,75	-0,41	1,00		
burro	-0,52	-0,34	-0,19	0,43	-0,46	0,29	0,10	1,00	
uova	-0,34	-0,31	-0,10	0,02	0,07	0,60	-0,22	0,45	1,00



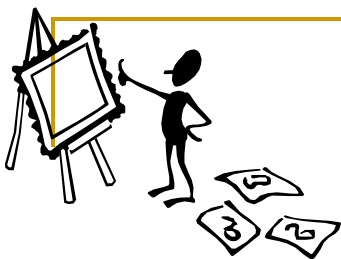
Analisi in R^p

(Spazio delle variabili)

$$\begin{aligned} \text{tr}(R) &= \sum_i r_{ii} \\ &= \sum_{\alpha} \lambda_{\alpha} \\ &(\alpha = 1, \dots, p) \end{aligned}$$

Matrice di correlazione:

	cereali	riso	patate	zucch.	verd.	carne	latte	burro	uova
cereali	1,00								
riso	0,13	1,00							
patate	0,06	0,25	1,00						
zucchero	-0,41	-0,69	-0,23	1,00					
verdure	0,56	0,57	0,07	-0,54	1,00				
carne	-0,07	-0,15	0,29	-0,19	0,22	1,00			
latte	-0,34	-0,39	-0,04	0,58	-0,75	-0,41	1,00		
burro	-0,52	-0,34	-0,19	0,43	-0,46	0,29	0,10	1,00	
uova	-0,34	-0,31	-0,10	0,02	0,07	0,60	-0,22	0,45	1,00



Analisi in R^p

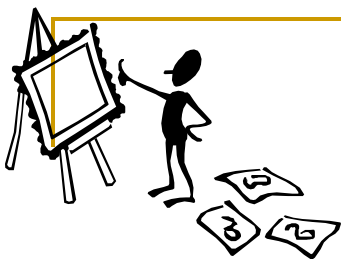
(Spazio delle variabili)

$$\begin{aligned} \text{tr}(R) &= \sum_i r_{ii} \\ &= \sum_{\alpha} \lambda_{\alpha} \\ &(\alpha = 1, \dots, p) \end{aligned}$$

Autovalori e percentuale di variabilità spiegata

Num	Valore	%	% cum.	Istogramma
1	3.4591	38.43	38.43
2	2.1630	24.02	62.47
3	1.1394	12.66	75.13
4	0.9345	10.38	85.51
5	0.4842	5.38	90.89
6	0.3251	3.61	94.50
7	0.2401	2.67	97.17
8	0.2143	2.38	99.55
9	0.0403	0.45	100.00

$$\frac{\lambda_{\alpha}}{\sum_{\alpha} \lambda_{\alpha}}$$



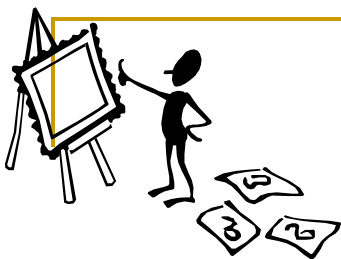
Analisi in R^p

(Spazio delle variabili)

Quante componenti scegliere ?

Autovalori e percentuale di variabilità spiegata

Num	Valore	%	% cum.	Istogramma
1	3.4591	38.43	38.43
2	2.1630	24.03	62.47
3	1.1394	12.66	75.13
4	0.9345	10.38	85.51
5	0.4842	5.38	90.89
6	0.3251	3.61	94.50
7	0.2401	2.67	97.17
8	0.2143	2.38	99.55
9	0.0403	0.45	100.00

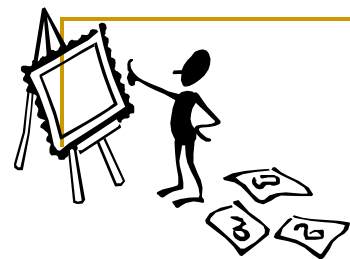


Analisi in R^p

(Spazio delle variabili)

Coordinate degli individui:

	(1)	(2)	(3)	(4)	(5)
Belgio	-0,39	2,06	-0,35	-0,19	-0,94
Danimarca	-1,73	1,20	0,64	0,90	0,85
Germania	-1,13	2,01	0,62	0,41	-0,16
Grecia	3,84	-0,64	0,95	0,65	-0,15
Spagna	2,70	1,48	-0,93	-0,45	1,87
Francia	-0,42	2,64	-0,01	-0,59	-0,78
Irlanda	-0,02	-0,19	-2,47	2,39	-0,25
Italia	2,92	-0,25	2,29	0,36	-0,46
Olanda	-0,61	-0,03	-0,89	-1,25	-0,35
Portogallo	2,22	-1,93	-1,05	-0,26	-0,80
Regno Unito	-0,29	-1,08	-0,60	-0,14	-0,17
Austria	-0,48	0,69	0,36	-0,48	0,43
Finlandia	-1,12	-1,62	-0,23	-1,57	-0,04
Islanda	-3,29	-1,53	1,31	1,26	-0,03
Norvegia	-0,92	-2,09	0,18	0,05	0,84
Svezia	-1,27	-0,72	0,16	-1,08	0,15



Analisi in R^n

(Spazio degli individui)

"Coordinate colonna"

	Asse1	Asse2	Asse3	Asse4	Asse5	Asse6	Asse7	Asse8	Asse9
Cereali	-0.6418	-0.2551	0.3203	0.5490	0.2187	0.2069	-0.1421	0.0629	-0.0651
Riso	-0.7193	-0.1622	-0.2347	-0.5967	0.0566	-0.0151	-0.0584	0.1672	-0.1018
Patate	-0.2897	0.0710	-0.8816	0.2869	0.0748	-0.0826	-0.1714	-0.0975	0.0125
Zucchero	0.8569	-0.1330	0.1441	0.1723	0.1065	-0.3993	-0.1139	0.0945	-0.0695
Verdure	-0.8815	0.2088	0.2548	-0.0055	-0.0631	-0.2143	-0.1576	0.1634	0.1137
Carne	-0.1134	0.8637	-0.1990	0.2871	0.0521	-0.0169	0.2561	0.2231	-0.0257
Latte	0.7066	-0.4998	-0.2510	0.0845	-0.2259	0.1950	-0.0829	0.2886	0.0389
Burro	0.6199	0.5186	0.0205	-0.2738	0.4729	0.1498	-0.1462	0.0423	0.0474
Uovo	0.1951	0.8605	0.1256	-0.0195	-0.3673	0.0956	-0.2355	-0.0487	-0.0581

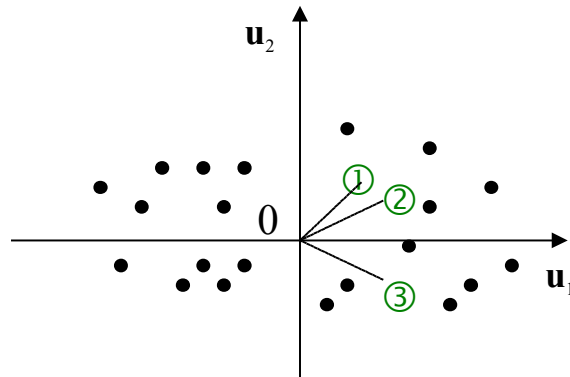
Analisi in Componenti Principali: Rappresentazione Simultanea

Coordinate delle variabili e
delle unità statistiche :

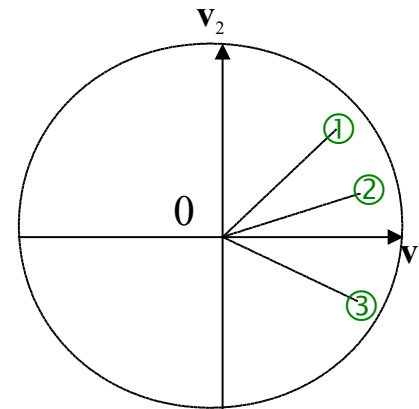
$$\hat{\phi}_{\alpha} = \sqrt{\lambda_{\alpha}} \mathbf{u}_{\alpha}$$

$$\hat{\psi}_{\alpha} = \sqrt{\lambda_{\alpha}} \mathbf{v}_{\alpha}$$

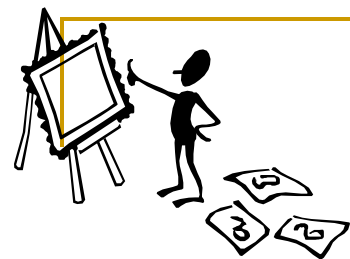
Spazio R^p delle unità statistiche



Spazio R^n delle variabili

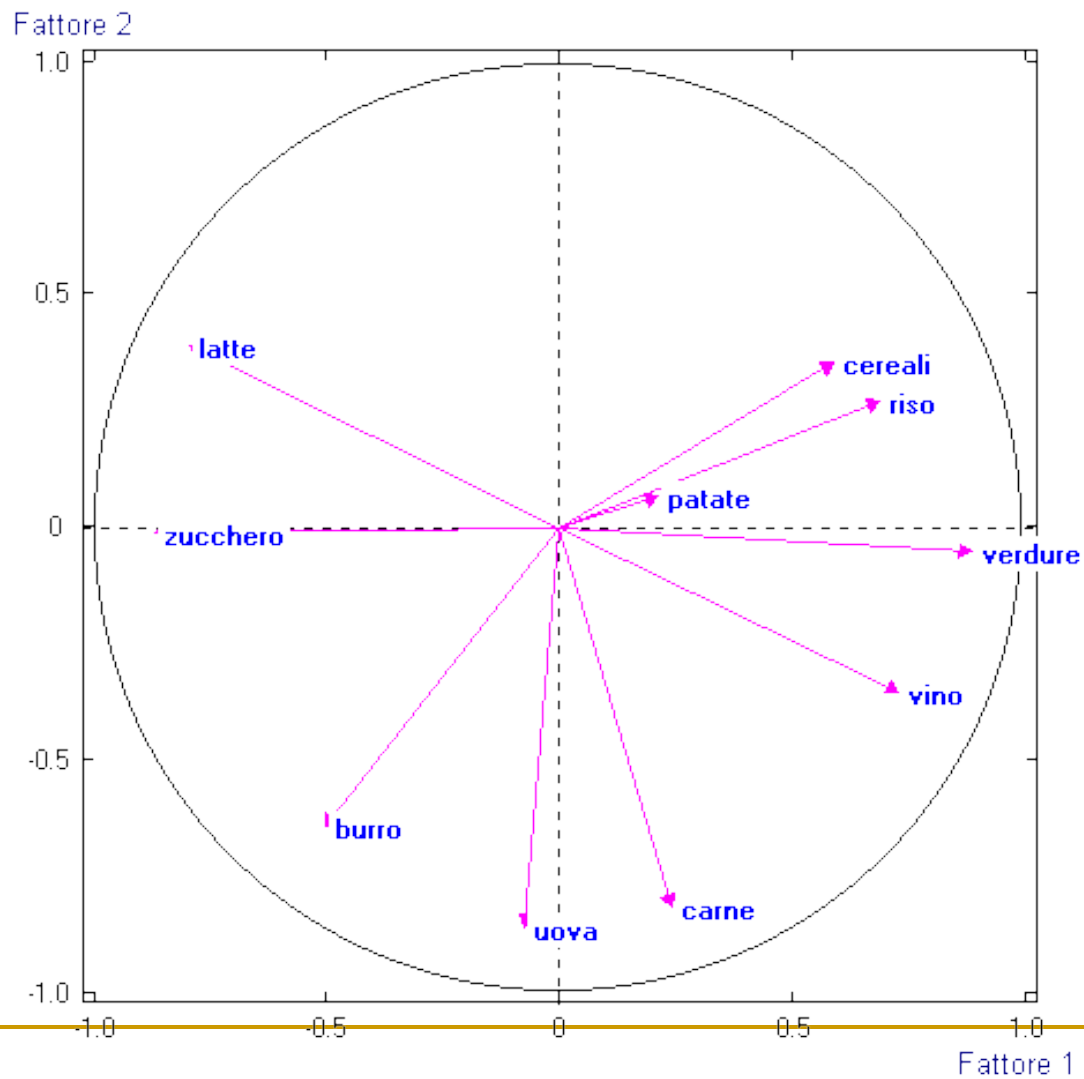


Le due nubi delle variabili coincidono (a meno di un fattore di scala)

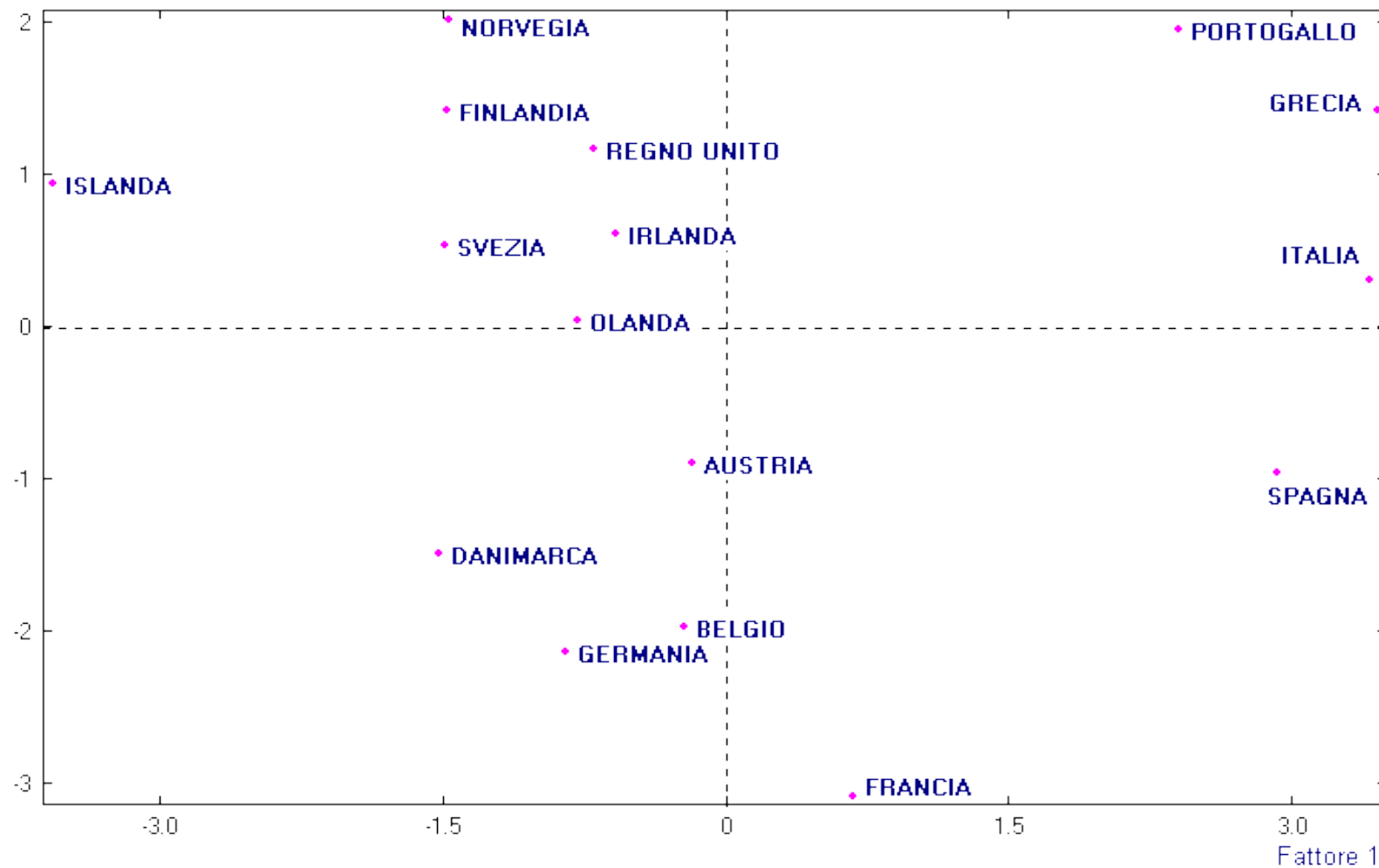


ACP

Rappresentazioni Grafiche



Fattore 2



Contributi assoluti e qualità della rappresentazione

Unità statistiche

	Dist	Dim.1	ctr	cos2	Dim.2	ctr	cos2	Dim.3	ctr	cos2
Belgio	2.478	-0.391	0.276	0.025	2.061	12.278	0.692	0.347	0.662	0.020
Danimarca	2.693	-1.733	5.428	0.414	1.203	4.182	0.199	-0.639	2.237	0.056
Germania	2.602	-1.127	2.296	0.188	2.006	11.627	0.594	-0.621	2.116	0.057
Grecia	4.165	3.842	26.676	0.851	-0.638	1.177	0.023	-0.951	4.960	0.052
Spagna	3.783	2.696	13.129	0.508	1.484	6.362	0.154	0.929	4.730	0.060
Francia	2.931	-0.416	0.312	0.020	2.638	20.115	0.811	0.008	0.000	0.000
Irlanda	3.539	-0.016	0.000	0.000	-0.188	0.102	0.003	2.468	33.408	0.486
Italia	3.852	2.919	15.395	0.574	-0.247	0.176	0.004	-2.293	28.839	0.354
Olanda	2.097	-0.612	0.676	0.085	-0.034	0.003	0.000	0.887	4.317	0.179
Portogallo	3.372	2.221	8.912	0.434	-1.930	10.759	0.327	1.046	6.005	0.096
RegnoUnito	1.502	-0.292	0.154	0.038	-1.080	3.372	0.517	0.597	1.958	0.158
Austria	1.243	-0.482	0.419	0.150	0.686	1.358	0.304	-0.356	0.694	0.082
Finlandia	2.822	-1.123	2.279	0.158	-1.622	7.598	0.330	0.228	0.285	0.007
Islanda	4.214	-3.293	19.595	0.611	-1.527	6.738	0.131	-1.314	9.477	0.097
Norvegia	2.567	-0.924	1.544	0.130	-2.093	12.654	0.665	-0.180	0.178	0.005
Svezia	2.084	-1.269	2.909	0.371	-0.720	1.498	0.119	-0.157	0.135	0.006

Contributi assoluti e qualità della rappresentazione

Variabili

	Dim.1	ctr	cos2	Dim.2	ctr	cos2	Dim.3	ctr	cos2
Cereali	0.642	11.907	0.412	-0.255	3.010	0.065	-0.320	9.005	0.103
Riso	0.719	14.957	0.517	-0.162	1.216	0.026	0.235	4.835	0.055
Patate	0.290	2.426	0.084	0.071	0.233	0.005	0.882	68.217	0.777
Zuccherio	-0.857	21.229	0.734	-0.133	0.818	0.018	-0.144	1.822	0.021
Verdure	0.882	22.465	0.777	0.209	2.016	0.044	-0.255	5.696	0.065
Carne	0.113	0.372	0.013	0.864	34.490	0.746	0.199	3.477	0.040
Latte	-0.707	14.435	0.499	-0.500	11.548	0.250	0.251	5.527	0.063
Burro	-0.620	11.109	0.384	0.519	12.433	0.269	-0.021	0.037	0.000
Uovo	-0.195	1.101	0.038	0.861	34.237	0.741	-0.126	1.384	0.016

Applicazione: ACP

	prezzo	cilindrata	cavalli	lungh.	largh.	peso	velocità	cons_strada	cons_urbano	affidab.
A155	29,9	1773	126	440	170	1325	200	8,4	9,5	136
AU80	34,2	1595	101	448	170	1270	178	8,8	10,5	408
BMW3	40,1	1796	115	443	170	1205	201	7,6	9,6	127
CXAN	29,4	1761	101	444	176	1176	188	7,1	10,3	118
TEMP	31,6	1756	101	435	170	1200	188	8,3	10,5	305
MOND	31,5	1796	111	448	175	1305	195	7,1	10,6	95
DELT	29,1	1756	103	401	170	1200	185	8,8	11	175
DEDR	31,1	1750	101	434	170	1255	180	8	10	184
PRIM	28,1	1597	102	446	170	1190	176	7,7	8,9	161
VECT	31	1796	90	435	170	1124	183	7,1	10,2	295
P405	28,5	1580	88	440	169	1080	173	8,1	9,8	280
RE21	29,5	1721	93	469	173	1075	185	8	10,8	350
GOLF	28,5	1781	90	434	170	1115	180	7,6	9,9	96
PASS	32	1781	90	461	172	1220	178	7,4	10,4	344
VOL4	32,4	1721	102	444	169	1086	180	7,4	10,3	148

Matrice delle Correlazioni

	prez	cili	cava	lung	larg	peso	velo	cons	cons	affi
prez	1.00									
cili	0.24	1.00								
cava	0.37	0.25	1.00							
lung	0.16	-0.14	-0.13	1.00						
larg	-0.11	0.33	0.05	0.38	1.00					
peso	0.25	0.20	0.70	-0.10	0.19	1.00				
velo	0.46	0.63	0.81	-0.04	0.30	0.49	1.00			
cons	-0.06	-0.41	0.18	-0.37	-0.48	0.23	-0.08	1.00		
cons	0.00	0.28	-0.25	-0.09	0.35	-0.07	0.00	0.15	1.00	
affi	0.06	-0.38	-0.49	0.35	-0.17	-0.16	-0.44	0.37	0.36	1.00

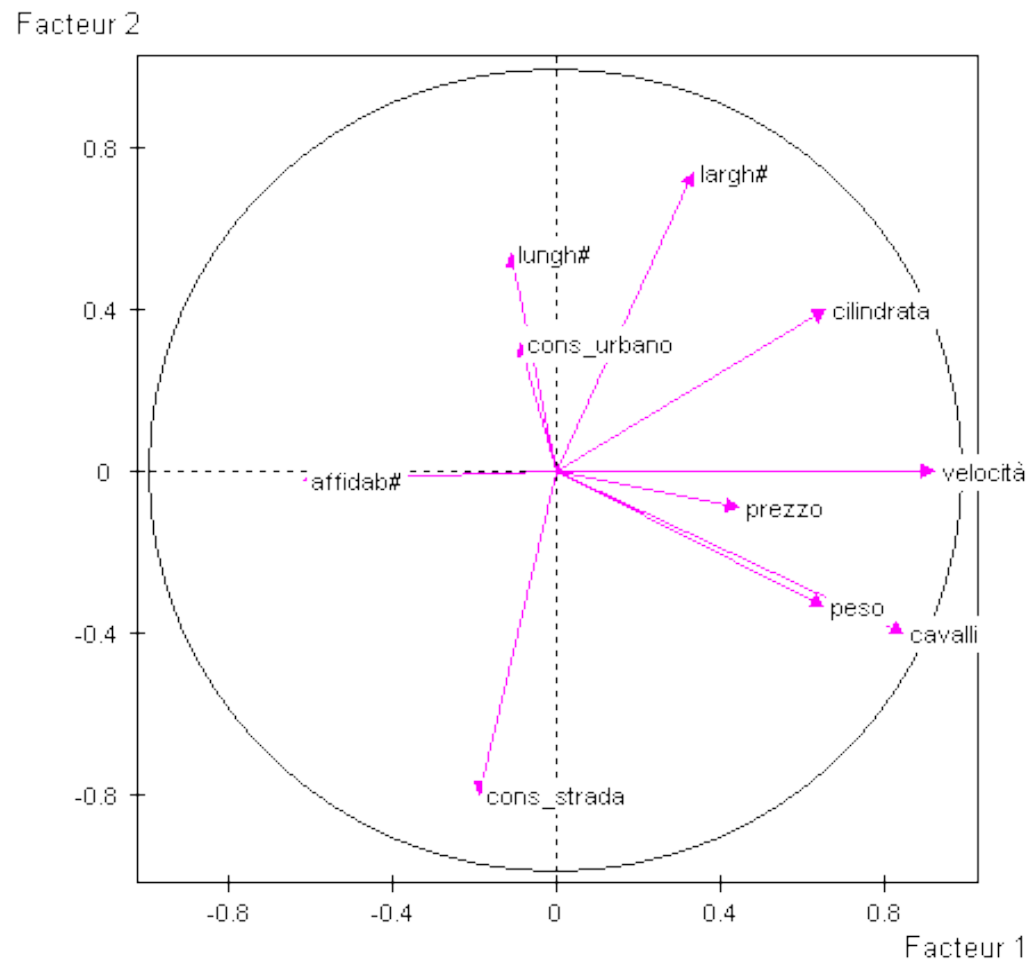
Autovalori
Istogramma dei primi 10 autovalori

NUMERO	VALEUR PROPRE	POURCENT.	POURCENT. CUMULE	
1	3.1844	31.84	31.84	*****
2	2.0321	20.32	52.17	*****
3	1.4892	14.89	67.06	*****
4	1.3389	13.39	80.45	*****
5	0.9852	9.85	90.30	*****
6	0.4000	4.00	94.30	*****
7	0.3348	3.35	97.65	*****
8	0.1319	1.32	98.97	****
9	0.0849	0.85	99.82	***
10	0.0184	0.18	100.00	*

Coordinate
VARIABLES ACTIVES

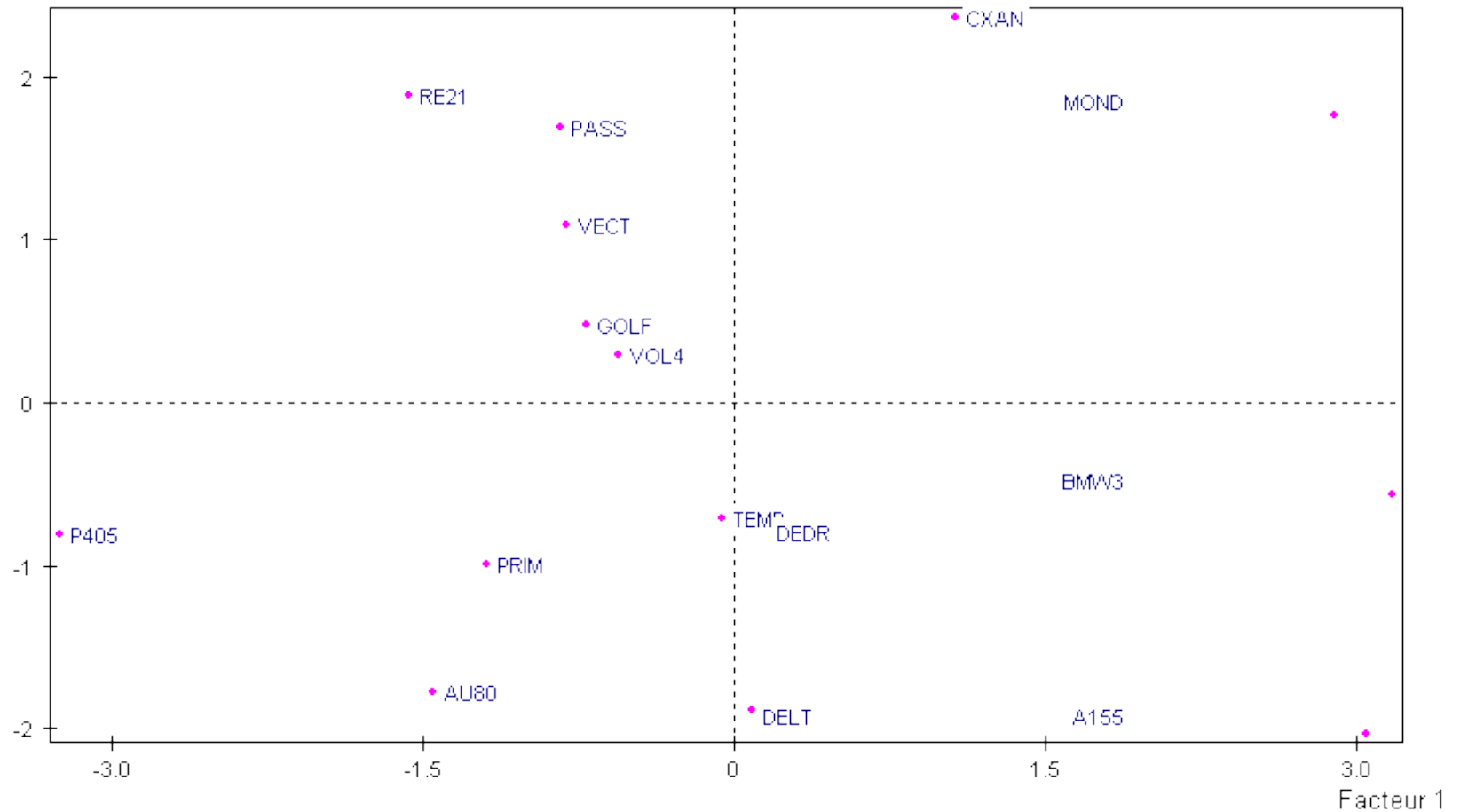
VARIABLES	COORDONNEES					CORRELATIONS VARIABLE-FACTEUR					ANCIENS AXES UNITAIRES				
IDEN - LIBELLE COURT	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
prez - prezzo	0.44	-0.09	0.38	0.43	-0.61	0.44	-0.09	0.38	0.43	-0.61	0.25	-0.06	0.31	0.38	-0.61
cili - cilindrata	0.66	0.40	0.06	-0.39	-0.33	0.66	0.40	0.06	-0.39	-0.33	0.37	0.28	0.05	-0.34	-0.34
cava - cavalli	0.85	-0.40	0.02	0.16	0.17	0.85	-0.40	0.02	0.16	0.17	0.48	-0.28	0.01	0.14	0.17
lung - lungh#	-0.11	0.54	0.21	0.74	0.15	-0.11	0.54	0.21	0.74	0.15	-0.06	0.38	0.17	0.64	0.15
larg - largh#	0.33	0.74	0.18	-0.07	0.49	0.33	0.74	0.18	-0.07	0.49	0.19	0.52	0.15	-0.06	0.49
peso - peso	0.65	-0.33	0.28	0.11	0.42	0.65	-0.33	0.28	0.11	0.42	0.37	-0.23	0.23	0.10	0.42
velo - velocità	0.92	0.00	0.14	0.01	-0.08	0.92	0.00	0.14	0.01	-0.08	0.52	0.00	0.12	0.01	-0.08
cons - cons_strada	-0.19	-0.80	0.43	-0.16	0.18	-0.19	-0.80	0.43	-0.16	0.18	-0.11	-0.56	0.35	-0.14	0.18
cons - cons_urbano	-0.09	0.32	0.71	-0.57	-0.03	-0.09	0.32	0.71	-0.57	-0.03	-0.05	0.22	0.58	-0.50	-0.03
affi - affidab#	-0.62	-0.02	0.69	0.22	-0.04	-0.62	-0.02	0.69	0.22	-0.04	-0.35	-0.02	0.57	0.19	-0.04

Variabili



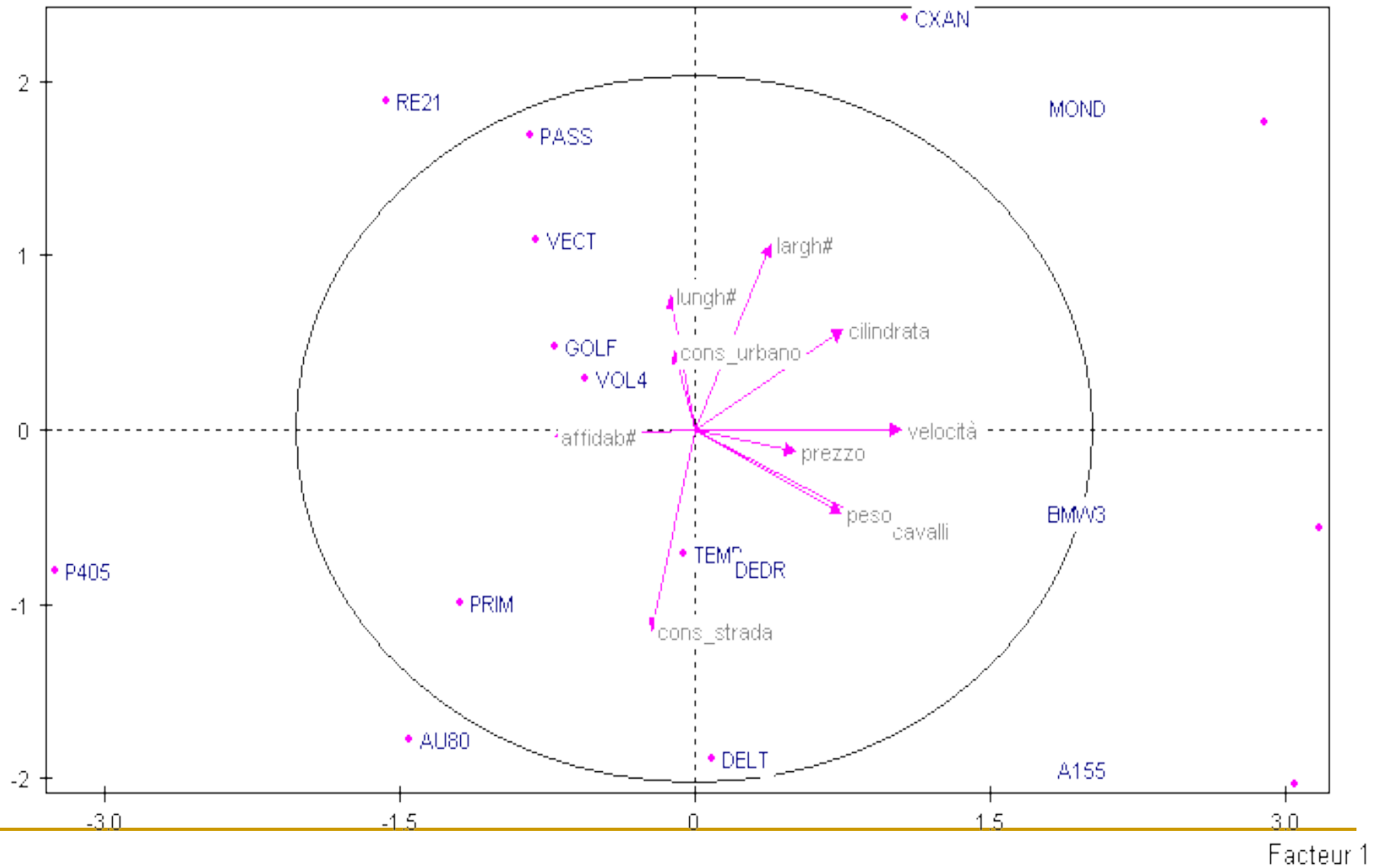
Piano degli individui

Facteur 2

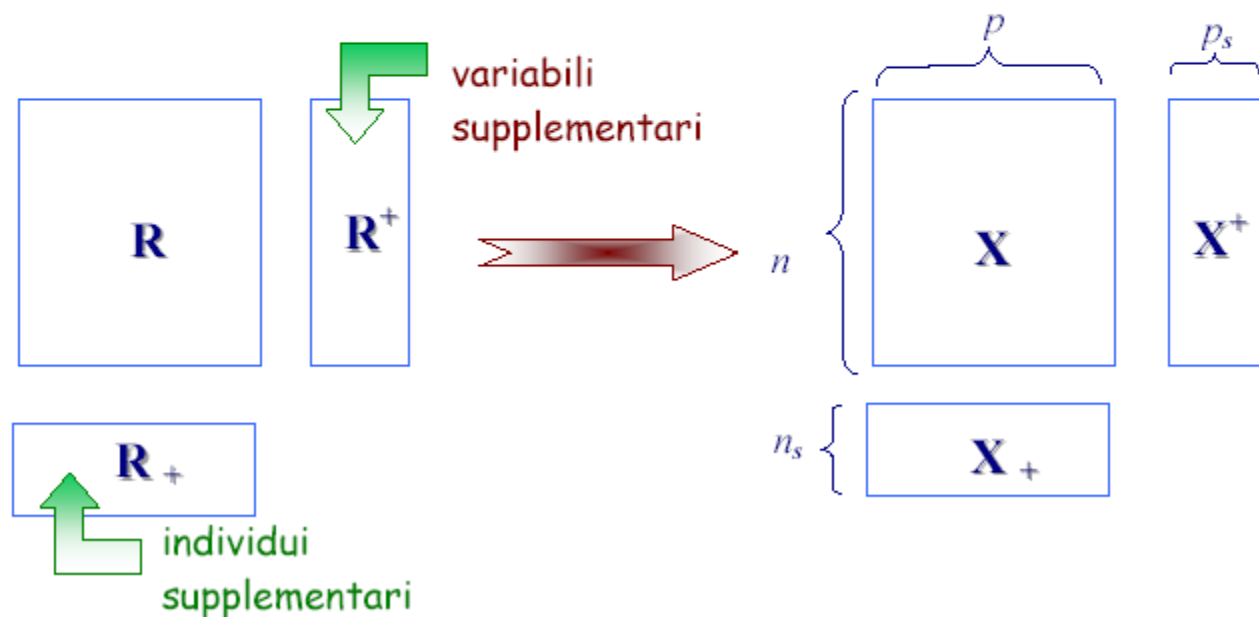


Rappresentazione simultanea

Facteur 2



Analisi in componenti principali: punti illustrativi



VARIABILI SUPPLEMENTARI (IN $|^N$)

- Trasformazioni: $x_{ij}^+ = \frac{1}{\sqrt{n}} \cdot \frac{r_{ij}^+ - \bar{r}_j^+}{s_j^+}$

- Coordinate su $\mathbf{v}_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} \mathbf{X} \mathbf{u}_\alpha$

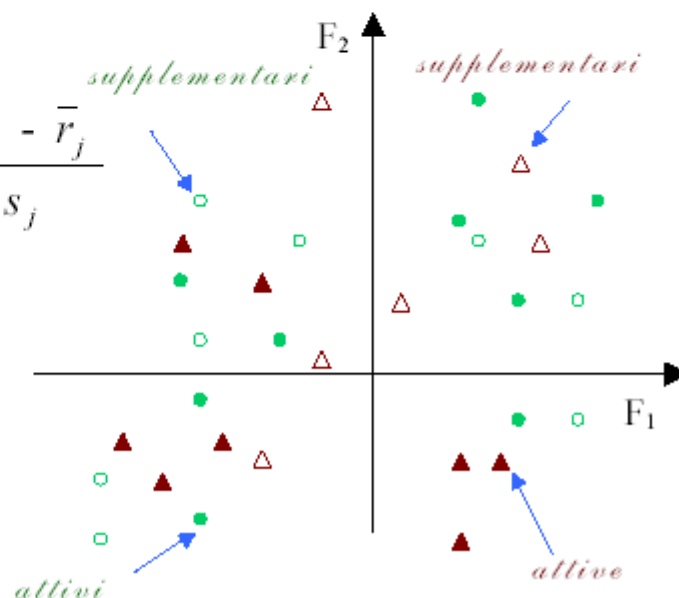
$$\hat{\phi}_\alpha^+ = (\mathbf{X}^+)' \mathbf{v}_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} (\mathbf{X}^+)' (\mathbf{X} \mathbf{u}_\alpha) = \frac{1}{\sqrt{\lambda_\alpha}} (\mathbf{X}^+)' \hat{\psi}_\varepsilon$$

INDIVIDUI SUPPLEMENTARI (IN $|^P$)

- Trasformazioni: $x_{+ij} = \frac{1}{\sqrt{n}} \cdot \frac{r_{+ij} - \bar{r}_j}{s_j}$

- Coordinate su \mathbf{u}_α

$$\hat{\psi}_\alpha^+ = \mathbf{X}_+ \mathbf{u}_\alpha$$



Proprietà di ottimalità delle componenti principali.

Notazioni

Sia \mathbf{X} , una matrice di dimensioni (n,p) ,

$0 < r \leq \text{rang}(\mathbf{X})$,

\mathbf{C} , la matrice delle prime r componenti principali di \mathbf{X} :

$$\mathbf{C} = \mathbf{X}\mathbf{U},$$

\mathbf{U} , la matrice dei primi r autovettori

- “Migliore approssimazione di una matrice di rango inferiore o uguale a r ”.

Le matrici \mathbf{C} e \mathbf{U} realizzano l’ottimo di:

$$\min_{(\mathbf{C}, \mathbf{U})} \|\mathbf{X} - \mathbf{C}'\mathbf{U}\|$$

- “Massimizzazione della norma”.

La matrice delle componenti principali \mathbf{C} realizza l’ottimo di

$$\max_{\mathbf{C}} \|\mathbf{C}\|$$

Proprietà di ottimalità delle componenti principali.

- “Massimizzazione del prodotto scalare”

Le matrici **C** e **U** realizzano l’ottimo di

$$\max_{(\mathbf{C}, \mathbf{U})} < \mathbf{X}, \mathbf{C}' \mathbf{U} >$$

- “Massimizzazione della proiezione”

Le matrici **C** e **U** realizzano l’ottimo di

$$\max_{(\mathbf{C}, \mathbf{U})} \left| \frac{\mathbf{X}, \mathbf{C}' \mathbf{U}}{\|\mathbf{C}' \mathbf{U}\|} \right|$$

- “Migliore approssimazione euclidea”

La matrice **C** è la migliore approssimazione euclidea della \mathbf{X} , essa realizza il

$$\min_{\mathbf{C}} (\min_{\mathbf{U}} \|\mathbf{X} - \mathbf{C}' \mathbf{U}\|)$$

Proprietà di ottimalità delle componenti principali.

- “Correlazione ottimale”

La matrice delle componenti principali \mathbf{C} realizza il

$$\max_{\mathbf{C}} \langle \mathbf{X}, \mathbf{Proj}_{\langle \mathbf{C} \rangle} \mathbf{X} \rangle$$

dove $\mathbf{Proj}_{\langle \mathbf{C} \rangle}$ denota la matrice del proiettore ortogonale sullo spazio generato dalle colonne di \mathbf{C}

- La matrice delle componenti principali \mathbf{C} realizza il

$$\max_{\mathbf{C}} \|\mathbf{X}'\mathbf{C}\|$$

- “Conservazione ottimale del prodotto scalare”

La matrice delle componenti principali \mathbf{C} realizza il

$$\min_{\mathbf{C}} \|\mathbf{XX}' - \mathbf{CC}'\|$$

Proprietà di ottimalità delle componenti principali.

- “Minimizzazione della matrice dei prodotti scalari dei residui”

Le matrici **C** e **U** realizzano l’ottimo di

$$\min_{(\mathbf{C}, \mathbf{U})} \|(\mathbf{X} - \mathbf{CU}')'(\mathbf{X} - \mathbf{CU}')\|$$

- “Massimizzazione simultanea degli autovalori”

$\forall i \leq r$ la matrice **C** realizza l’ottimo di

$$\max_{\mathbf{C}} \lambda_i(\mathbf{C}'\mathbf{C})$$

- “Minimizzazione simultanea degli autovalori”

$\forall i \leq r$ le matrici **C** e **U** realizzano l’ottimo di

$$\min_{(\mathbf{C}, \mathbf{U})} \lambda_i[(\mathbf{X} - \mathbf{C}'\mathbf{U})'(\mathbf{X} - \mathbf{C}'\mathbf{U})]$$