

# Analisi Fattoriale

## Concetti introduttivi

Marcello Gallucci

Milano-Bicocca

A  
M  
D

- L'Analisi Fattoriale (e varianti) si propone di estrarre un numero limitato di fattori (variabili latenti o sottostanti) da un set di variabili osservate (e.s. items), al fine di rappresentare al meglio la variabilità di tale set
- Ciò consente di interpretare le relazioni tra un gran numero di variabili osservate mediate un numero limitato di fattori

# Indicatori vs costruito

Il costruito è una dimensione latente non osservabile

Tratto latente

Estroversione

Si diverte molto

Ama chiacchierare

Ama le feste

Prende l'iniziativa

Indicatore: misura empirica osservabile (comportamento, item etc.)

# Indicatori

- ⇒ Un buon indicatore deve essere una buona misura del costrutto e **non** deve essere una buona misura di altri costrutti
- ⇒ Più indicatori sono necessari per definire un costrutto
- ⇒ La definizione concreta (*operazionalizzazione*) di un costrutto deriva dagli indicatori
- ⇒ Molto spesso in Psicologia il costrutto è ipotetico/latente mentre gli indicatori sono osservabili/empirici

# Costrutto

- ⇒ Un costrutto puo' essere definito come un un concetto (dimensione, fattore, tratto, classe, componente) teorico con certe conseguenze empiriche
- ⇒ Gli indicatori misurano le conseguenze empiriche
- ⇒ Il costrutto associa gli indicatori (legno puo' associare tavolo, sedia, quadro, bastone, albero, casa, barca)
- ⇒ Da un punto di vista statistico, se alcuni indicatori correlano tra di essi, possiamo inferire la presenza di un costrutto sottostante (o viceversa)

# Due tipi di indicatori e costrutti

## Indicatori riflessivi

Estroversione

Loquace

Vivace

Dinamico

Assumiamo che siano correlati

## Indicatori formativi

Alcolismo

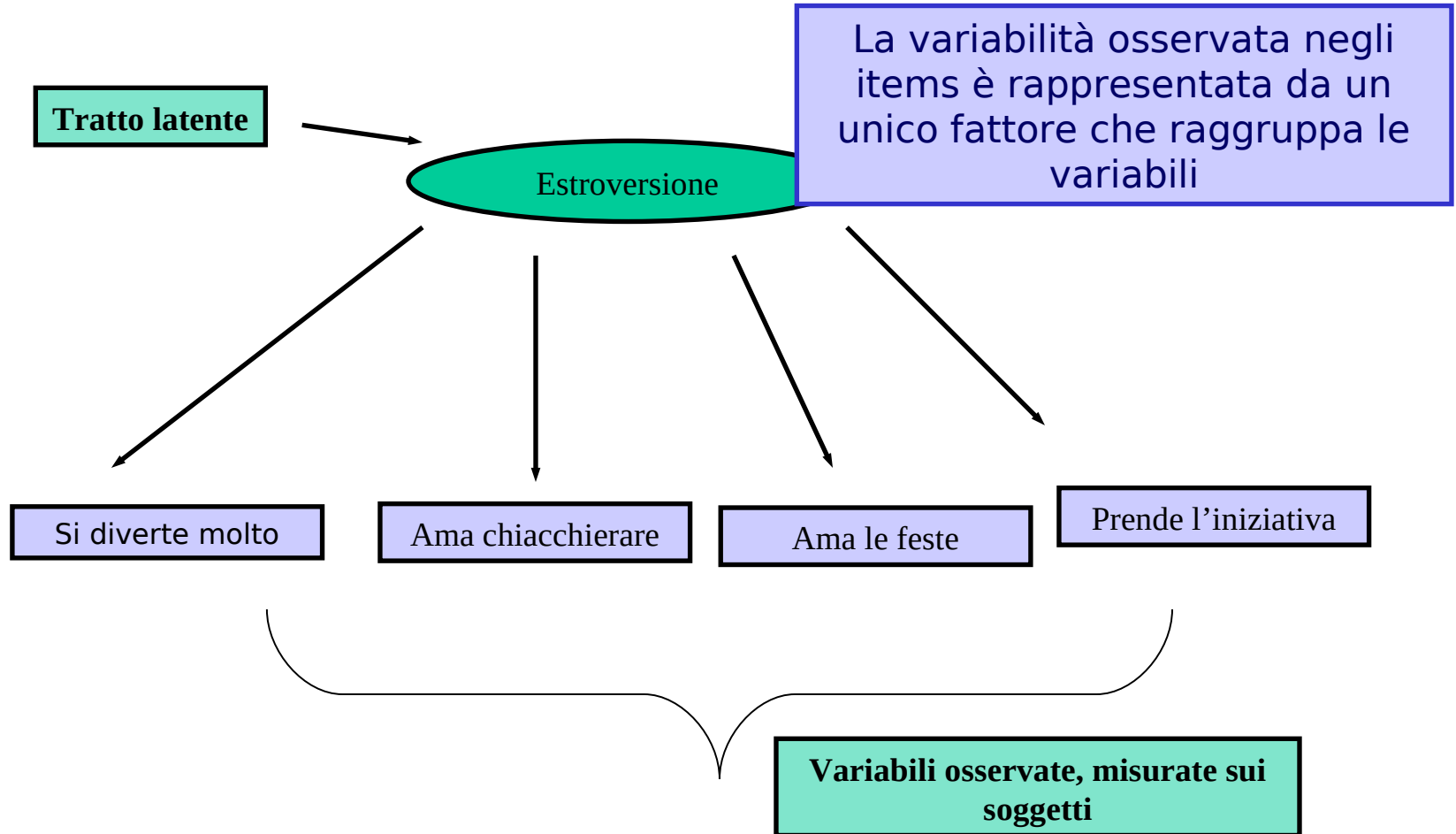
Whisky

Vino

Birra

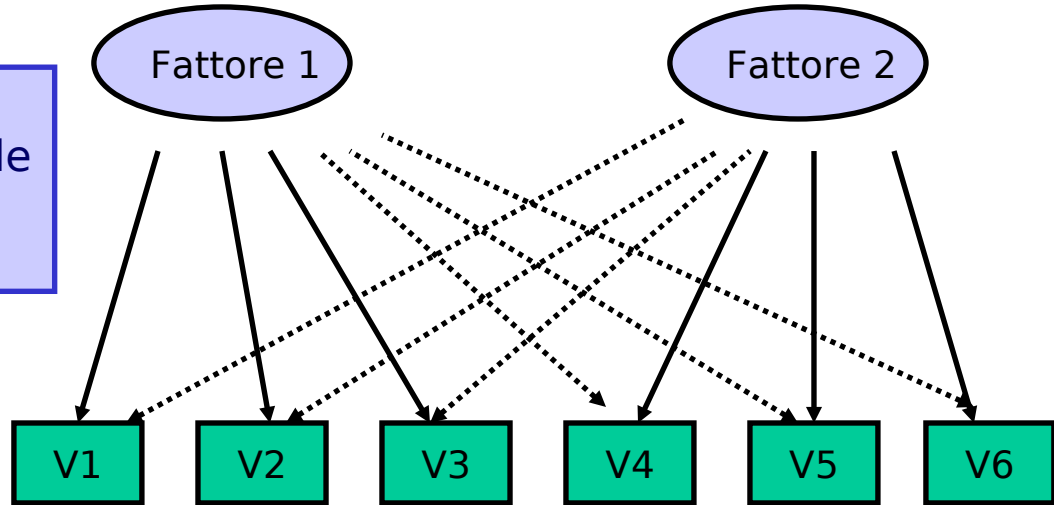
Non assumiamo che siano correlati

# Un possibile modello



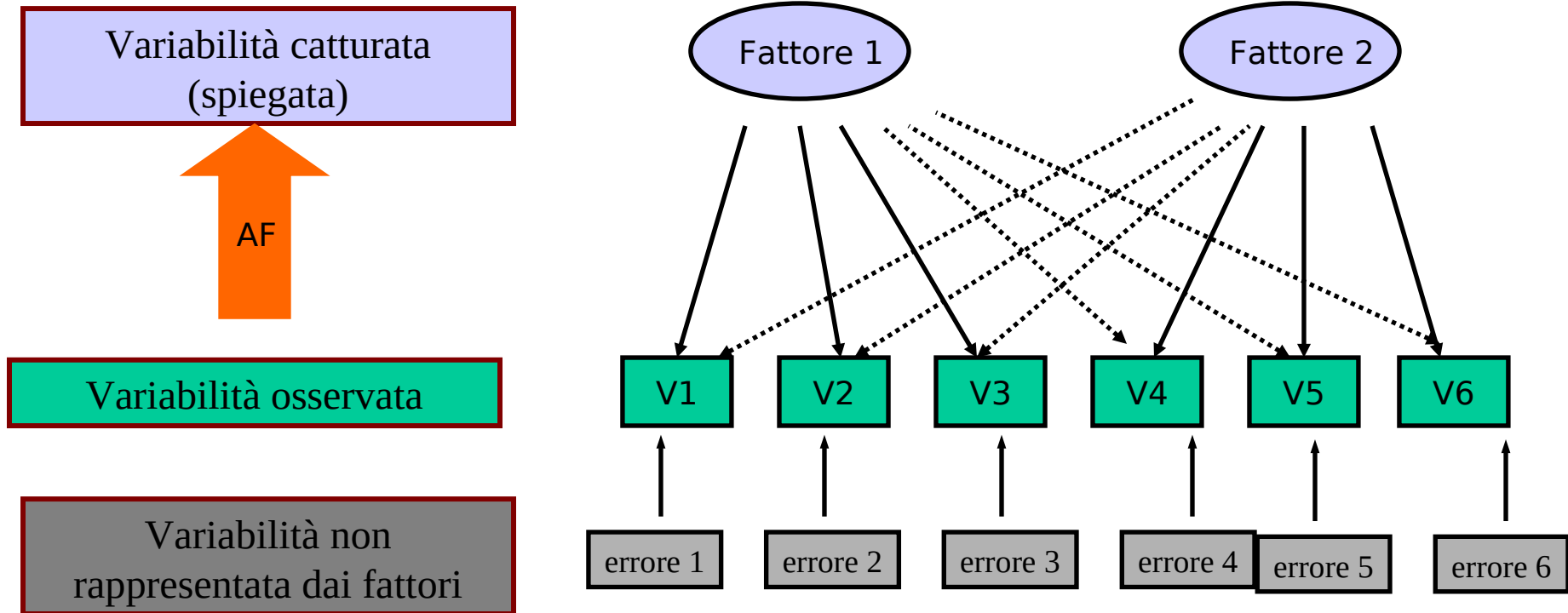
# Un altro possibile modello

Ovviamente i fattori utili a rappresentare la variabilità delle variabili osservate possono essere numerosi

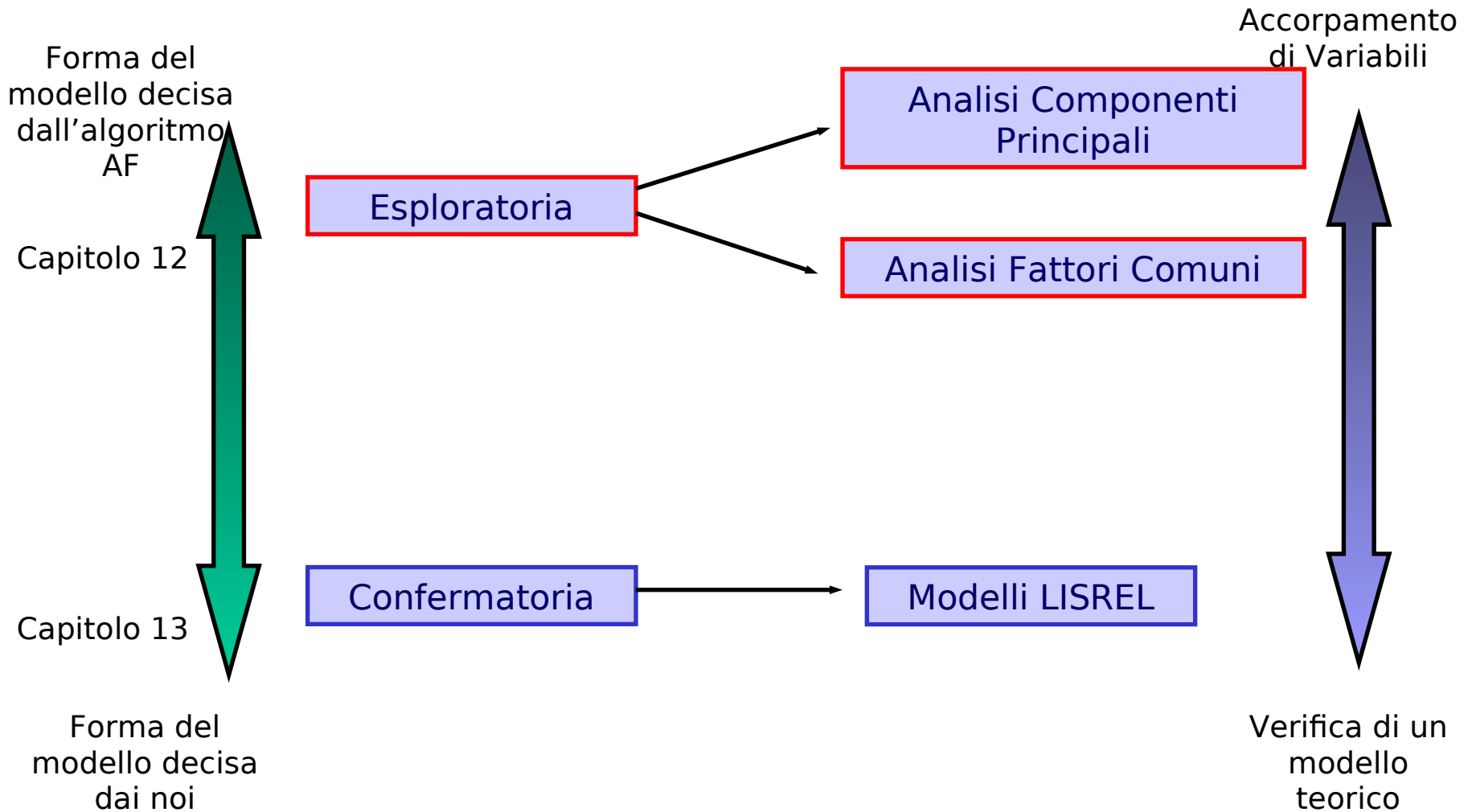




Guadagnando in parsimonia ed interpretabilità delle relazioni, non ci si può aspettare di non perdere qualcosa in precisione



# Tipi di Analisi Fattoriale



- Alcuni concetti sono utili per ogni tipo di Analisi Fattoriale, indipendentemente dalle differenze tecniche (che vedremo successivamente) tra questi tipi di analisi
- Tutte le varianti dell'AF: Rappresentazione delle **relazioni** fra variabili misurate mediante un numero ristretto di fattori

- Le relazioni fra variabili (continue) sono calcolabili mediante il coefficiente  $r$  di correlazione di Pearson.

## Matrice di correlazione

Correlazioni

		a1	a2	a3	a4
a1	Correlazione di Pearson	1	.084	.154	.242*
	Sig. (2-code)		.409	.126	.015
	N	100	100	100	100
a2	Correlazione di Pearson	.084	1	.514**	.231*
	Sig. (2-code)	.409		.000	.021
	N	100	100	100	100
a3	Correlazione di Pearson	.154	.514**	1	.588**
	Sig. (2-code)	.126	.000		.000
	N	100	100	100	100
a4	Correlazione di Pearson	.242*	.231*	.588**	1
	Sig. (2-code)	.015	.021	.000	
	N	100	100	100	100

\*. La correlazione è significativa al livello 0,05 (2-code).

\*\*.. La correlazione è significativa al livello 0,01 (2-code).

A1: Si diverte molto

A2: Ama chiacchierare

A3: Ama le feste

A4: Prende l'iniziativa

- Un altro modo per definire lo scopo dell'Analisi Fattoriale è l'estrazione di un numero ristretto di fattori che riproducano al meglio la matrice di correlazione osservata

Matrice di correlazione

Correlazioni

		a1	a2	a3	a4
a1	Correlazione di Pearson	1	.084	.154	.242*
	Sig. (2-code)		.409	.126	.015
	N	100	100	100	100
a2	Correlazione di Pearson	.084	1	.514**	.231*
	Sig. (2-code)	.409		.000	.021
	N	100	100	100	100
a3	Correlazione di Pearson	.154	.514**	1	.588**
	Sig. (2-code)	.126	.000		.000
	N	100	100	100	100
a4	Correlazione di Pearson	.242*	.231*	.588**	1
	Sig. (2-code)	.015	.021	.000	
	N	100	100	100	100

\*. La correlazione è significativa al livello 0,05 (2-code).

\*\*.. La correlazione è significativa al livello 0,01 (2-code).

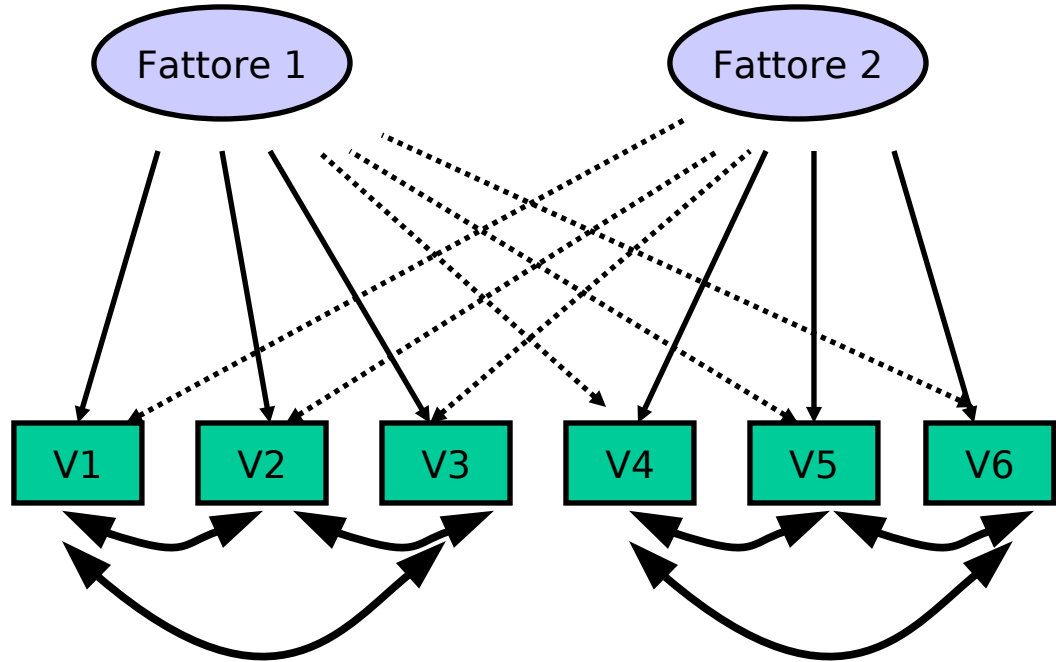
A1: Si diverte molto

A2: Ama chiacchierare

A3: Ama le feste

A4: Prende l'iniziativa

# Relazioni fra variabili e fattori latenti



L'idea di fondo è che le variabili osservate correlino perché condividono un fattore sottostante

# Relazioni fra variabili e fattori latenti

Dunque vogliamo creare delle nuove variabili (fattori) che combinino le variabili osservate che correlano molto

Correlazioni

		a1	a2	a3	a4
a1	Correlazione di Pearson	1	.084	.154	.242*
	Sig. (2-code)		.409	.126	.015
	N	100	100	100	100
a2	Correlazione di Pearson	.084	1	.514**	.231*
	Sig. (2-code)	.409		.000	.021
	N	100	100	100	100
a3	Correlazione di Pearson	.154	.514**	1	.588**
	Sig. (2-code)	.126	.000		.000
	N	100	100	100	100
a4	Correlazione di Pearson	.242*	.231*	.588**	1
	Sig. (2-code)	.015	.021	.000	
	N	100	100	100	100

E separino le variabili che non correlano fra loro

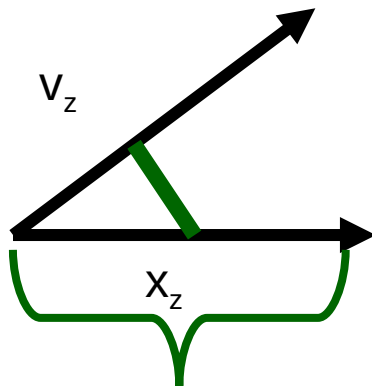
\*. La correlazione è significativa al livello 0,05 (2-code).

\*\* . La correlazione è significativa al livello 0,01 (2-code).

# Rappresentazione vettoriale

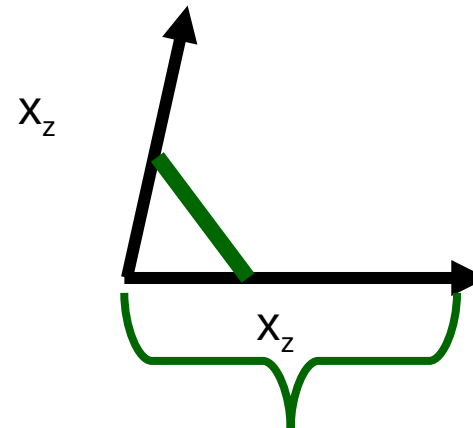
- Ogni variabile può essere rappresentata mediante un vettore di lunghezza uguale alla sua deviazione standard
- L'associazione tra due variabili si può rappresentare mediante l'angolo tra i due vettori (prodotto tra i vettori = prodotto tra gli  $z$  = correlazione)

Angolo acuto =  
associazione forte



Lunghezza = std.dev = 1

Angolo meno acuto =  
Associazione debole

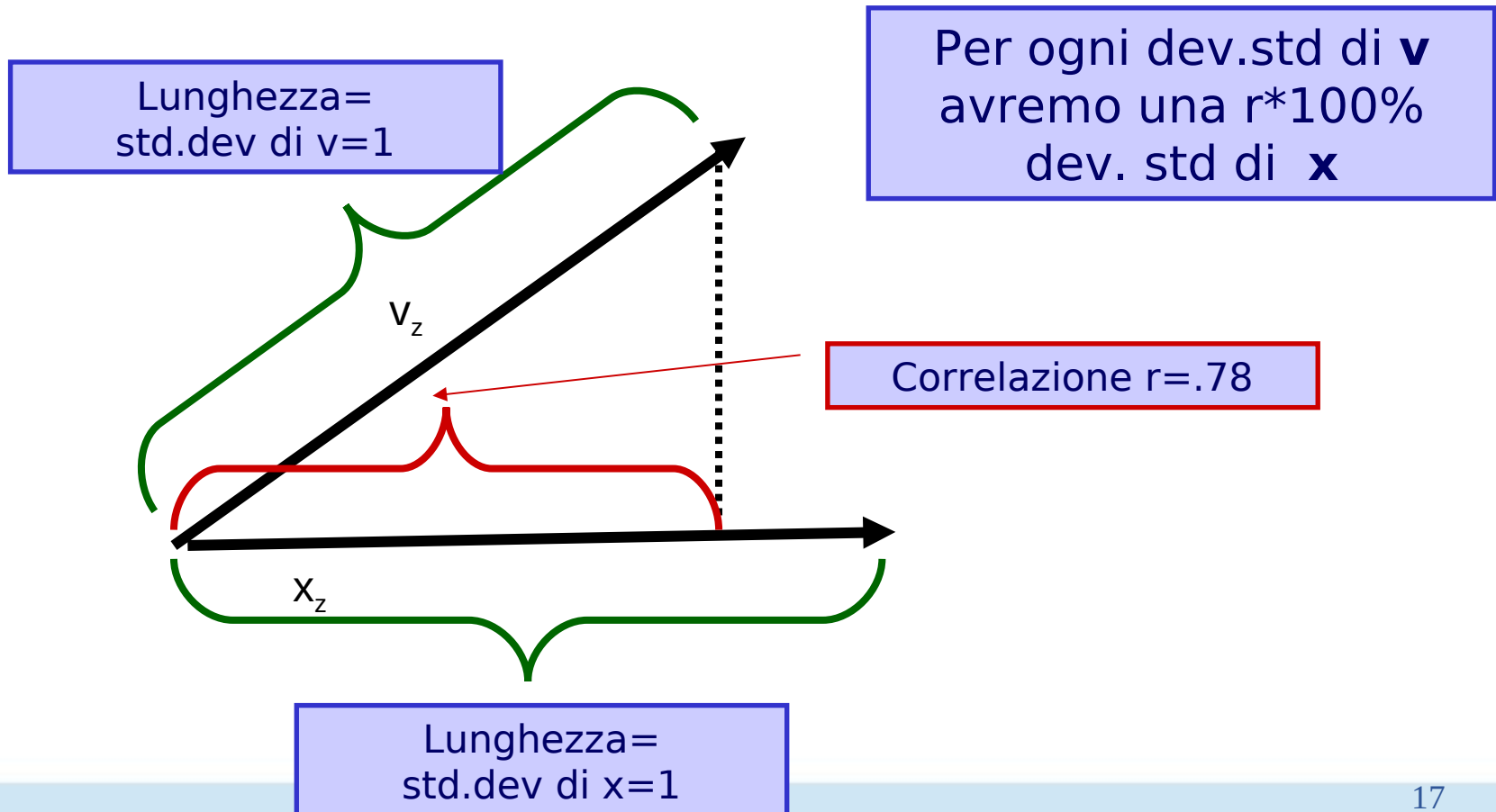


Lunghezza = std.dev = 1



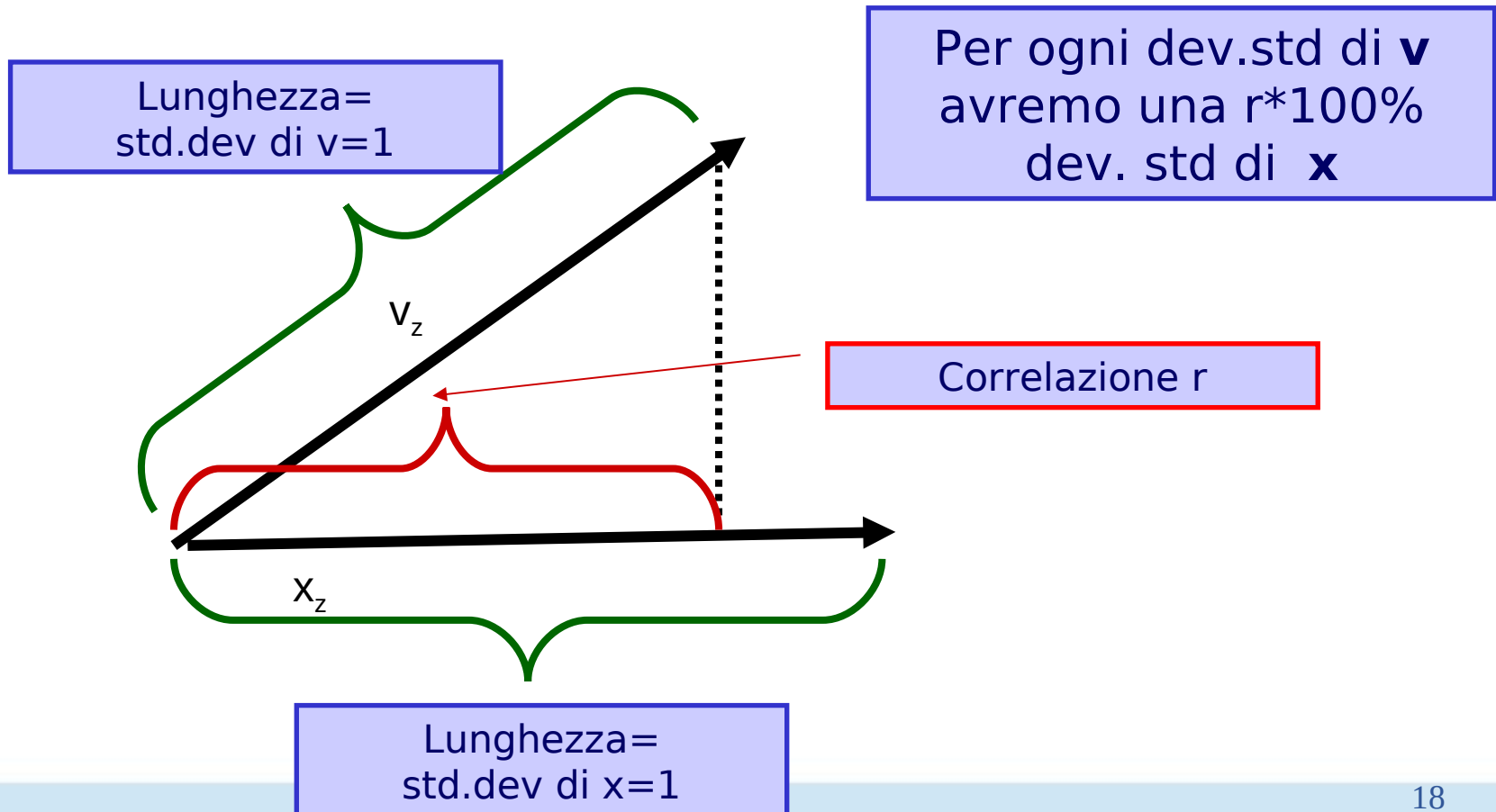
# Rappresentazione vettoriale

- La correlazione quantifica la proiezione di un vettore-variabile sull'altro!



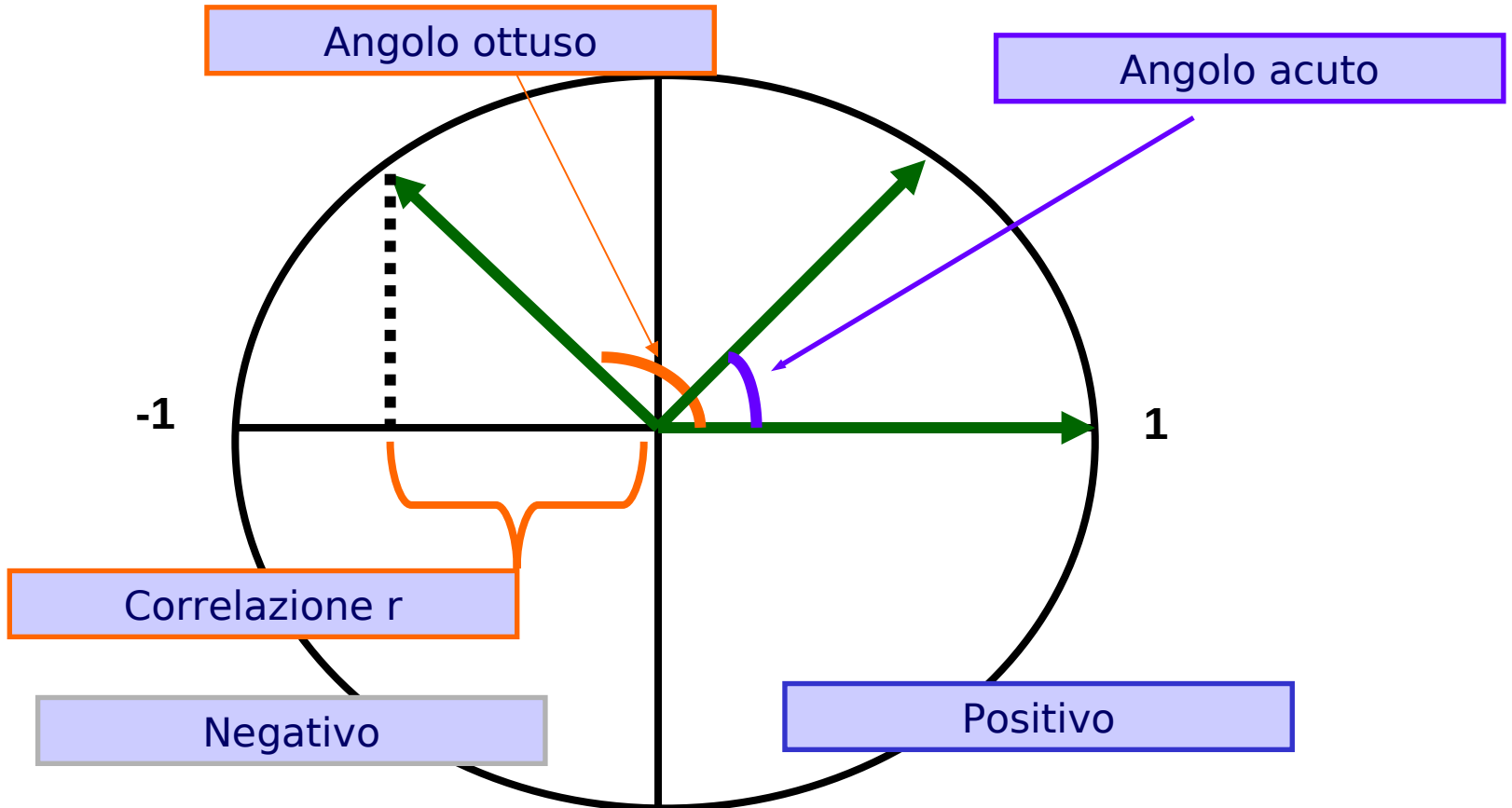
# Rappresentazione vettoriale

- La correlazione quantifica la **proiezione** di un vettore-variabile sull'altro!



# Rappresentazione vettoriale

- correlazione positiva o negativa

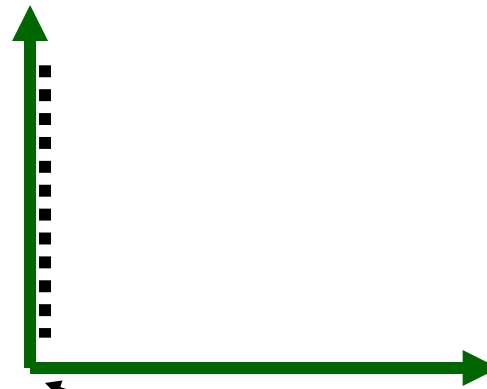


# Rappresentazione vettoriale

- Se la proiezione è zero, cioè  $\mathbf{r}=\mathbf{0}$ , le due variabili saranno indipendenti (linearmente)

Questo è il motivo per cui spesso si dice che due variabili non correlate sono **ortogonali**

E che due variabili correlate sono **oblique**

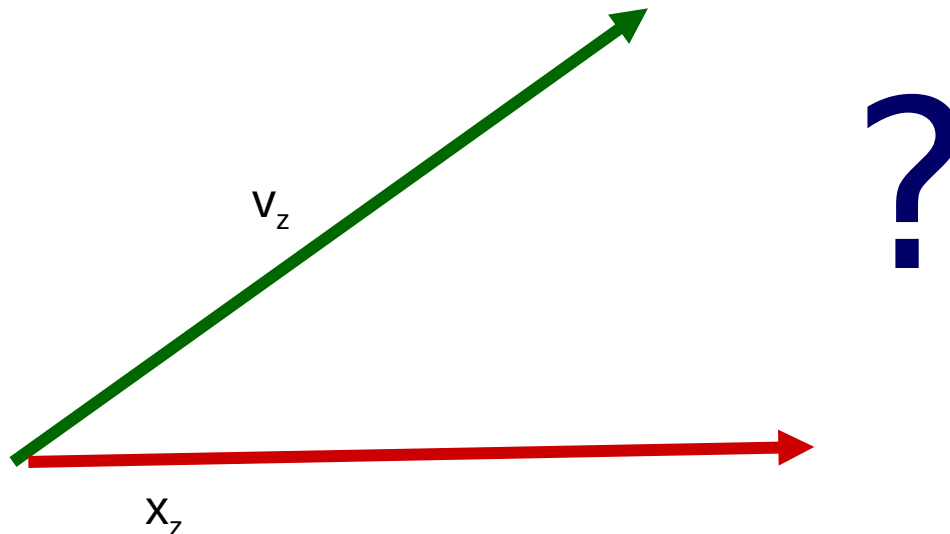


Un vettore non proietta nulla sull'altro

- In tutte le varianti dell'AF il fine è di estrarre una serie di fattori che siano al “centro” dell'insieme di variabili

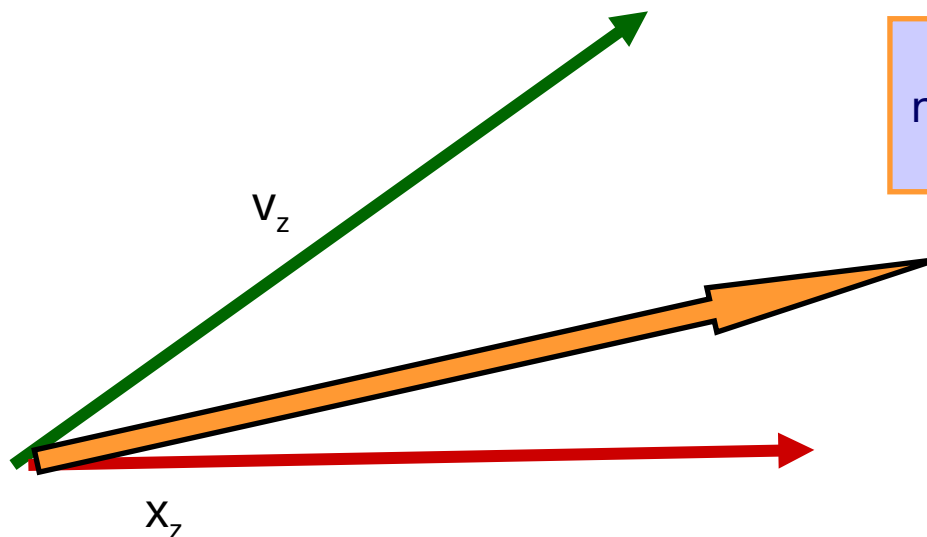
# Estrazione Fattore Comune

- Applichiamo la rappresentazione vettoriale: Ci proponiamo di rappresentare le due variabili qui sotto mediante un fattore unico: Dove sarà questo fattore?



# Estrazione Fattore Comune

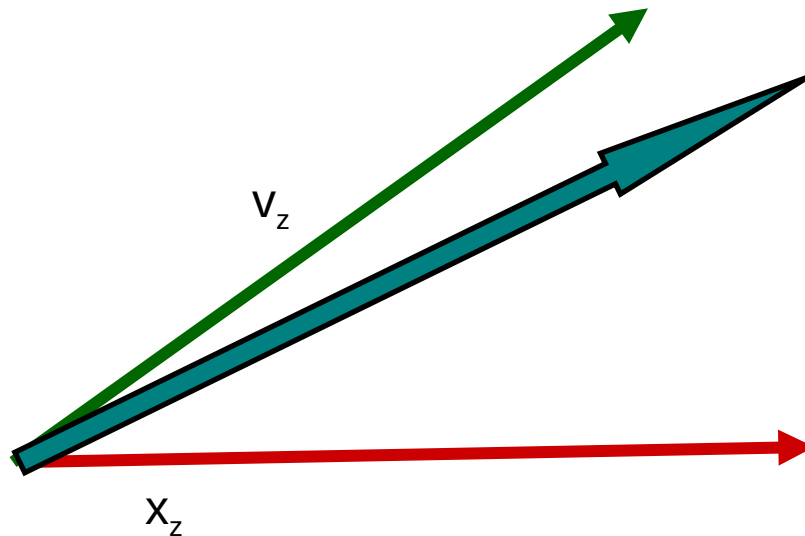
- Il fattore deve essere una nuova variabile che meglio rappresenti entrambe le variabili



In questa posizione  
rappresenterebbe bene  
X ma non V

# Estrazione Fattore Comune

- Il fattore deve essere una nuova variabile che meglio rappresenti entrambe le variabili

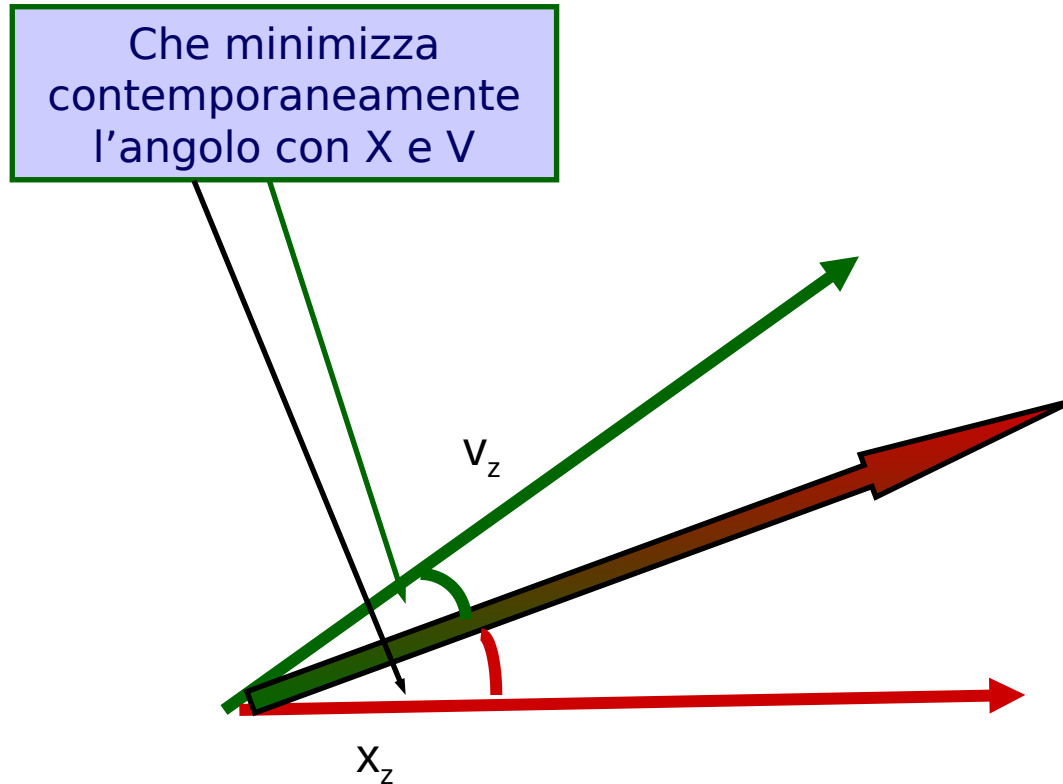


In quest'altra troppo bene V ma non X



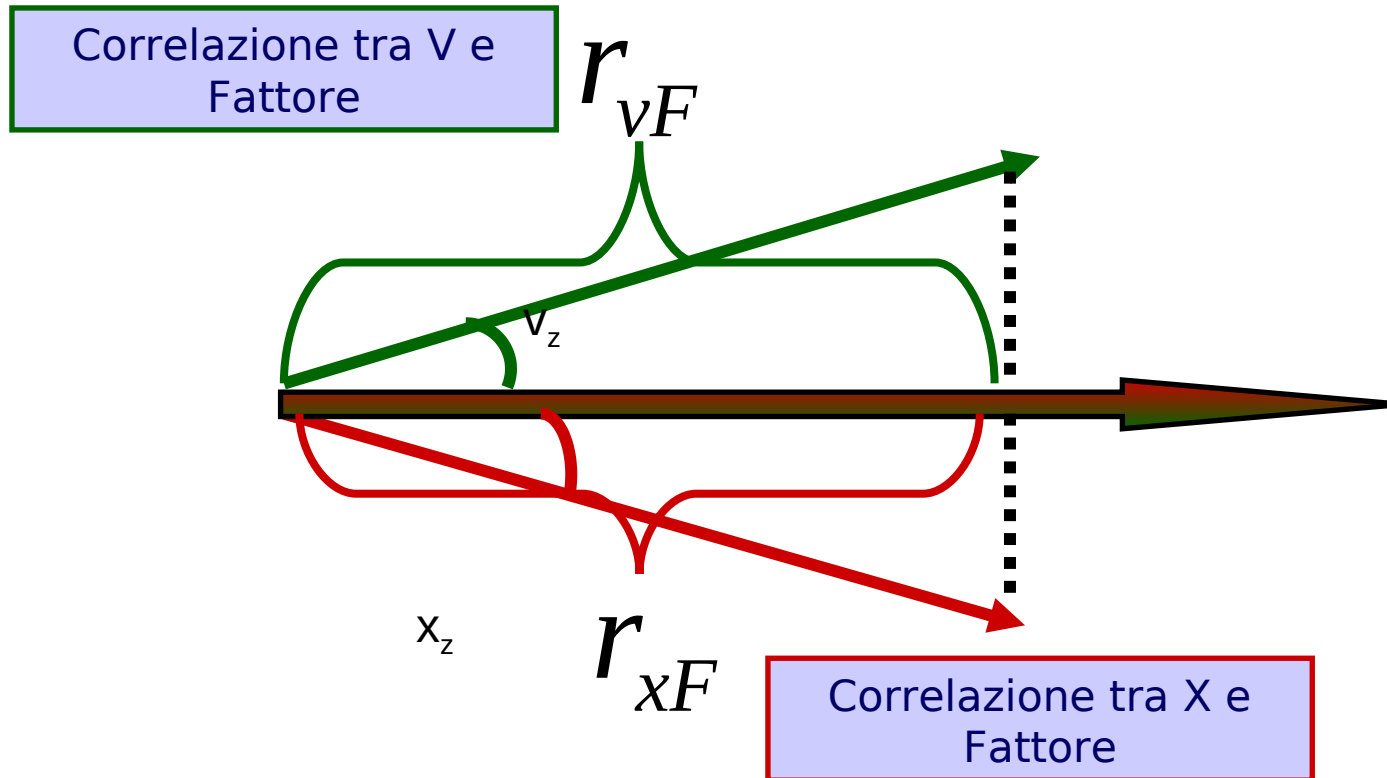
# Estrazione Fattore Comune

- Il fattore comune sarà al centro! Che vuol dire precisamente?



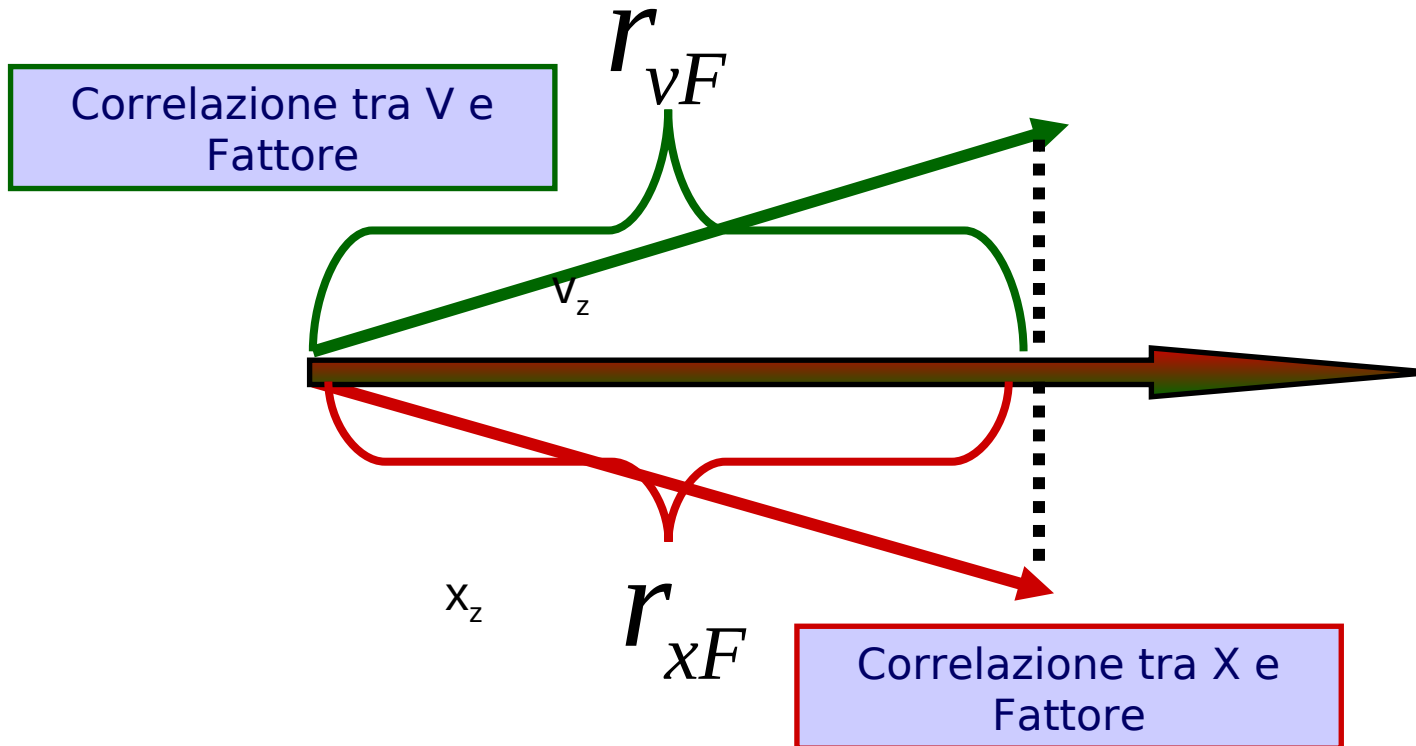
# Estrazione Fattore Comune

- L'angolo è tanto più piccolo quanto più è alta la correlazione



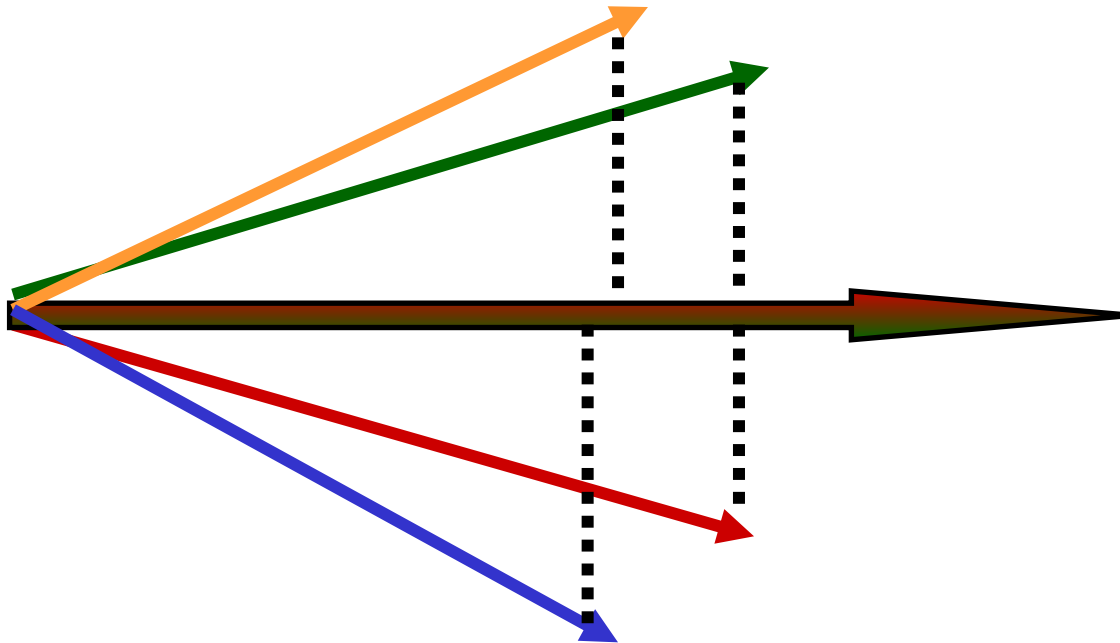
# Estrazione Fattore Comune

- L'angolo è tanto più piccolo quanto più è alta la correlazione
- **Dunque il miglior fattore è quello che massimizza le correlazioni con le variabili osservate**



# Le correlazioni possono variare

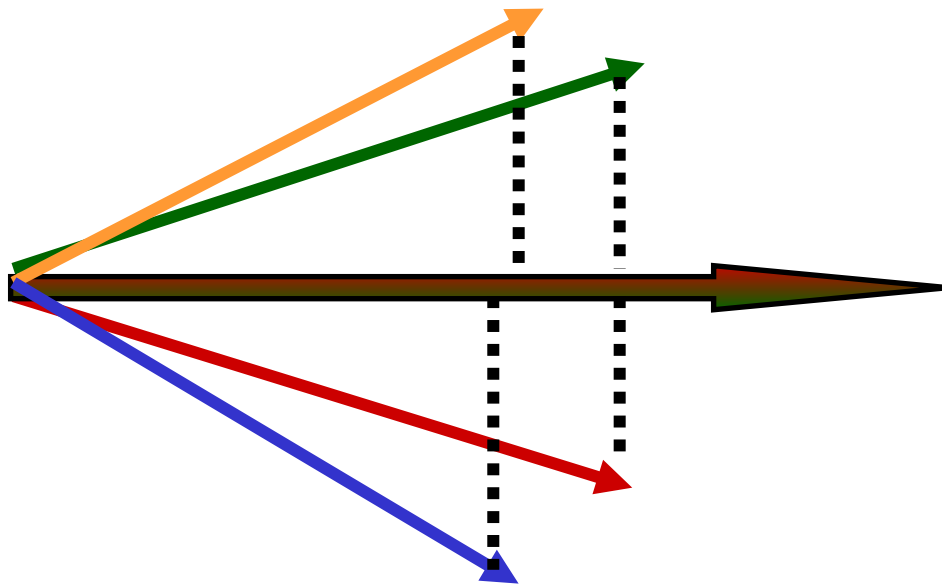
- Nel caso generale (non due variabili) le correlazioni con il fattore sono differenti per le varie variabili



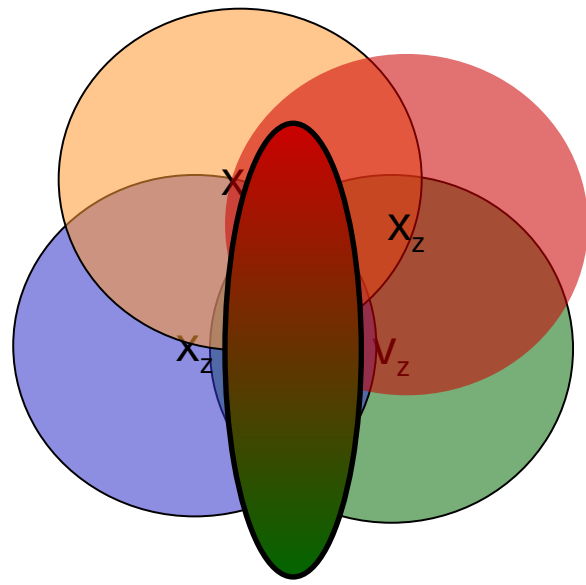
# Correlazioni e Varianze

- Ricorda che la correlazione (al quadrato) indica la varianza condivisa

Correlazioni



Varianze

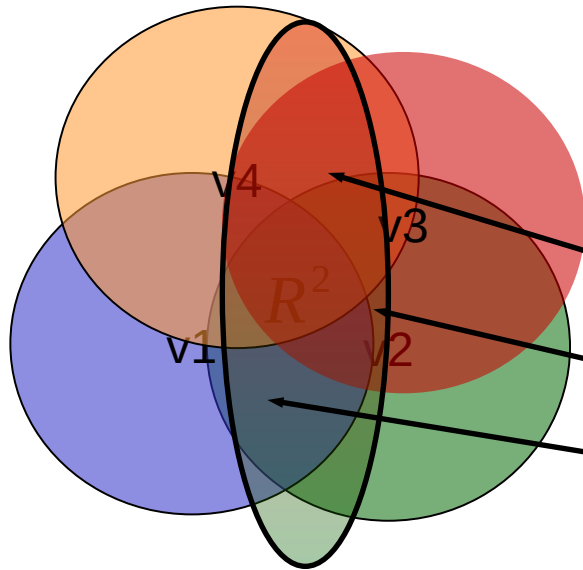


Dunque il miglior fattore è quello che meglio cattura la varianza condivisa

# Varianza spiegata dal fattore

- Quanto sarà questa varianza?

Varianze



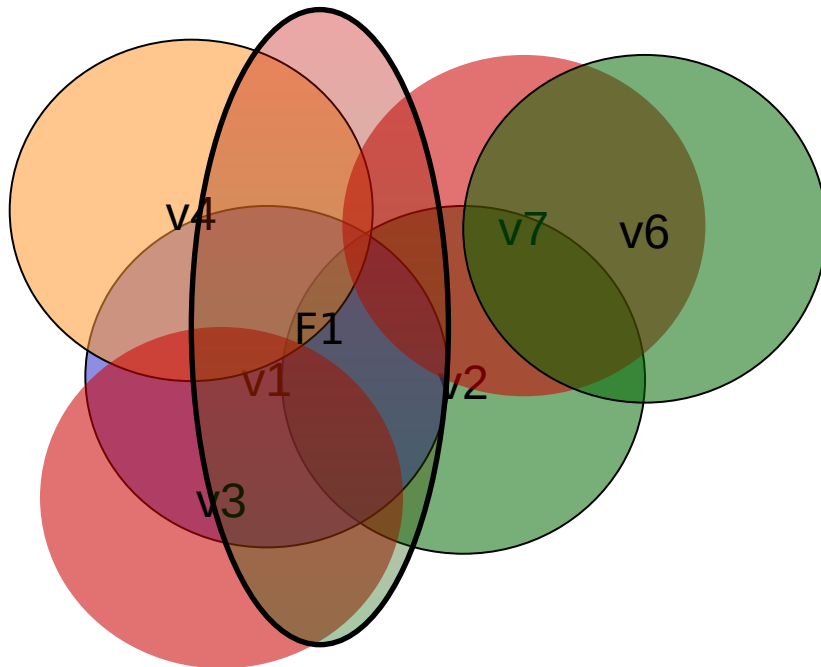
La varianza spiegata dal fattore sarà la somma delle varianze che condivide con ogni singola variabile

Dunque sarà la somma dei quadrati delle correlazioni tra variabili e fattori

$$\text{var}(F) = r_{v1F}^2 + r_{v2F}^2 + r_{v3F}^2 + \dots$$

- Consideriamo di aver estratto un fattore da questo insieme di variabili

Varianze

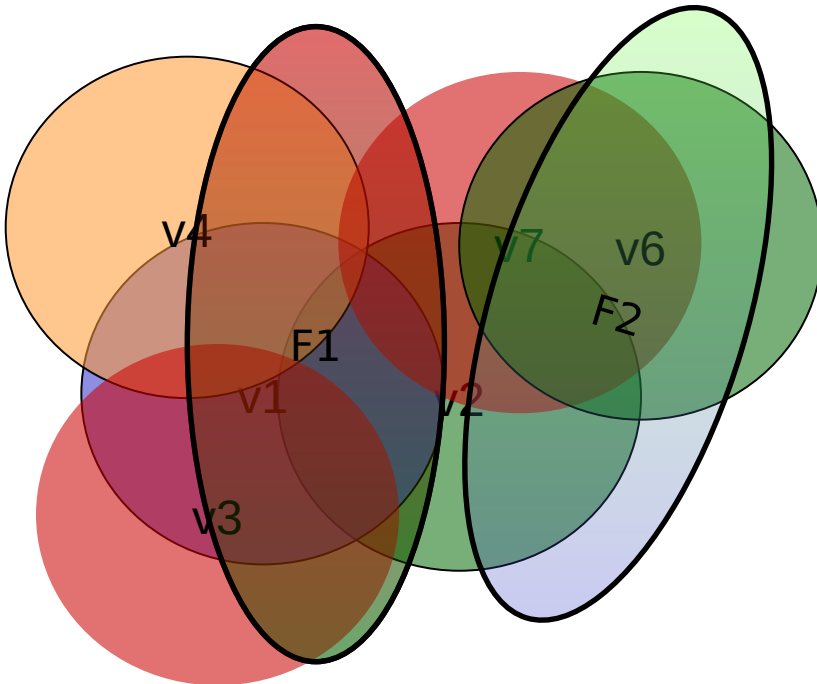


Il fattore che estraiamo sarà quello che massimizza la varianza spiegata

Ma non necessariamente cattura tutta la varianza condivisa

- Avremo così rappresentato le varianze osservate mediante due fattori

Varianze

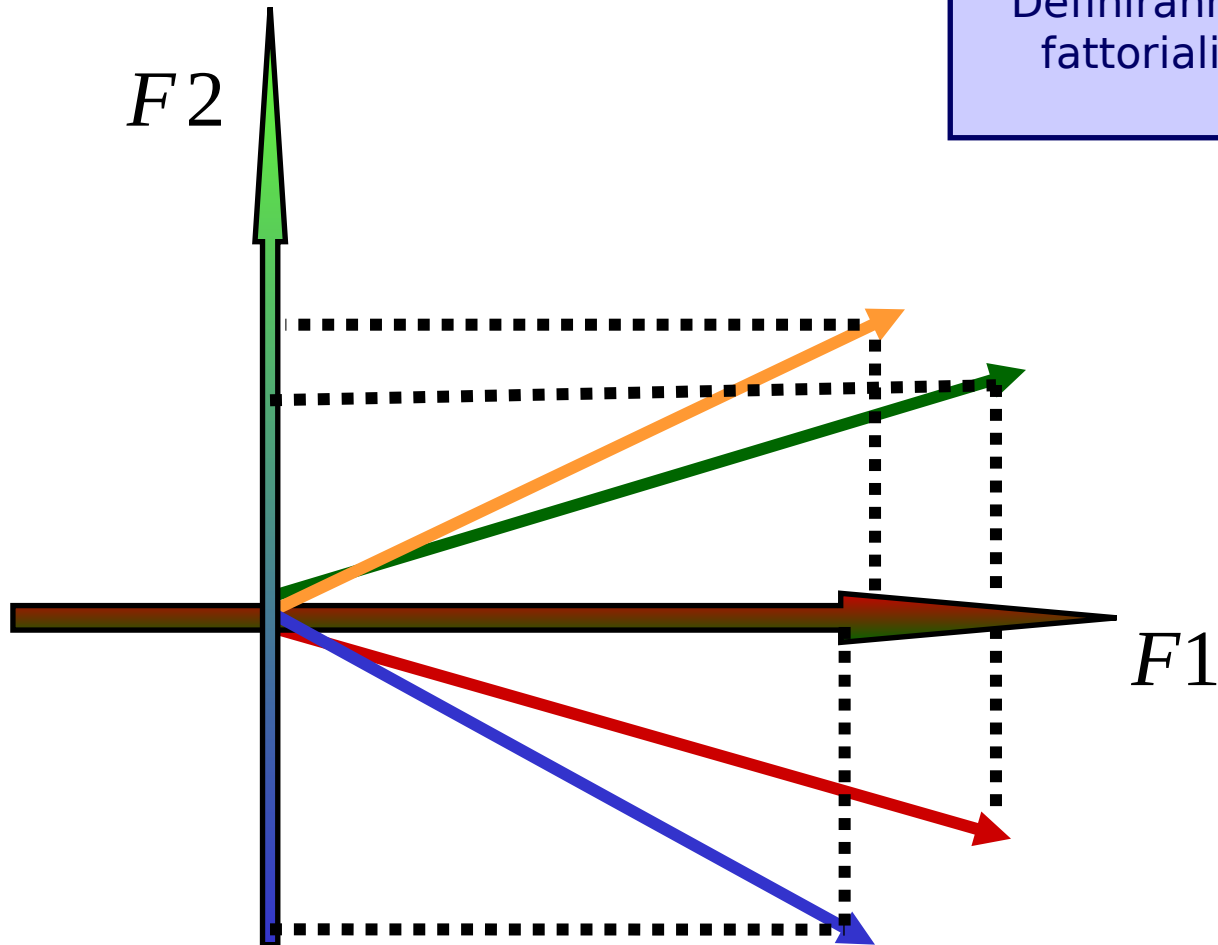


I fattori non condividono  
varianza, dunque non sono  
correlati



# Estrazione di più fattori

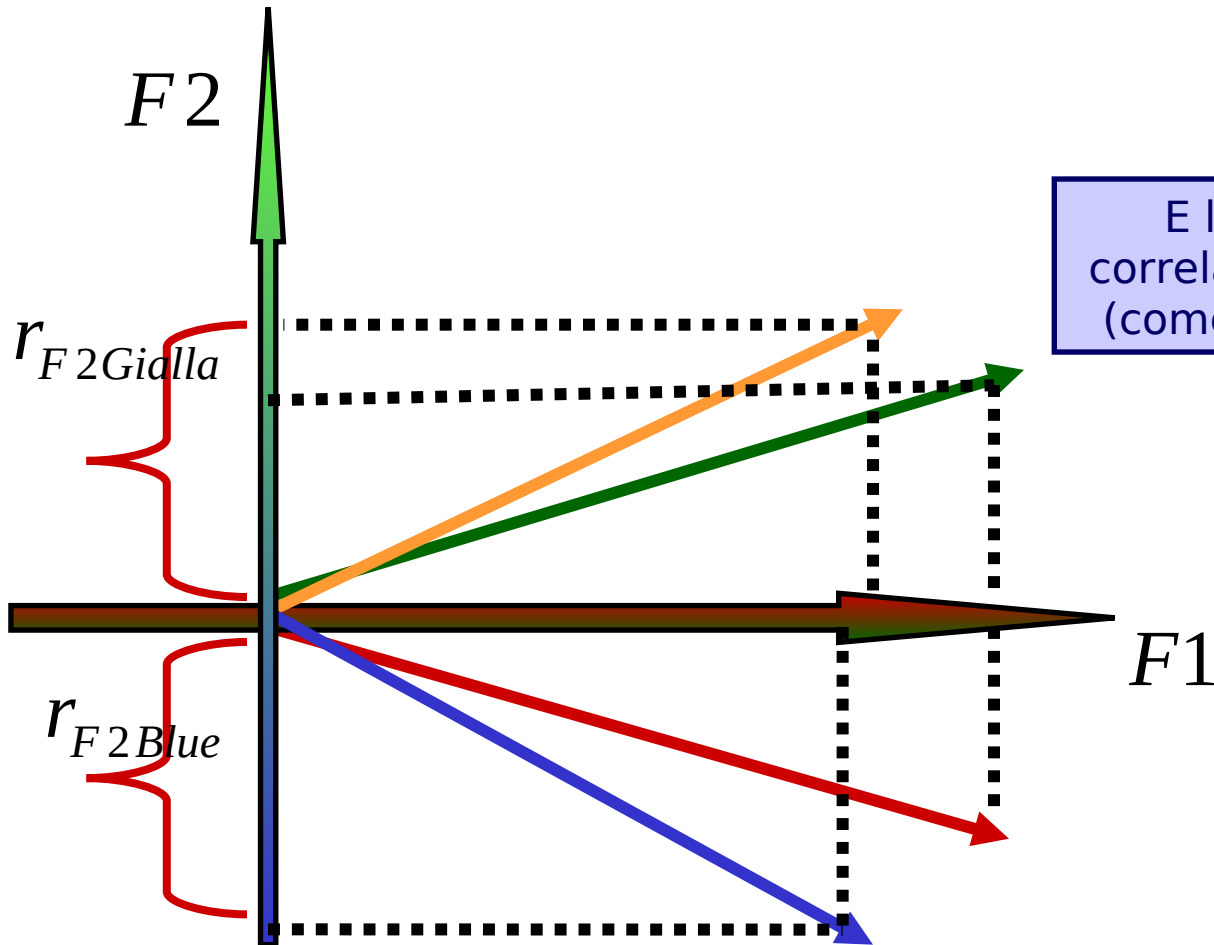
- Dunque saranno ortogonali



Definiranno dunque degli assi fattoriali dove proiettare le variabili

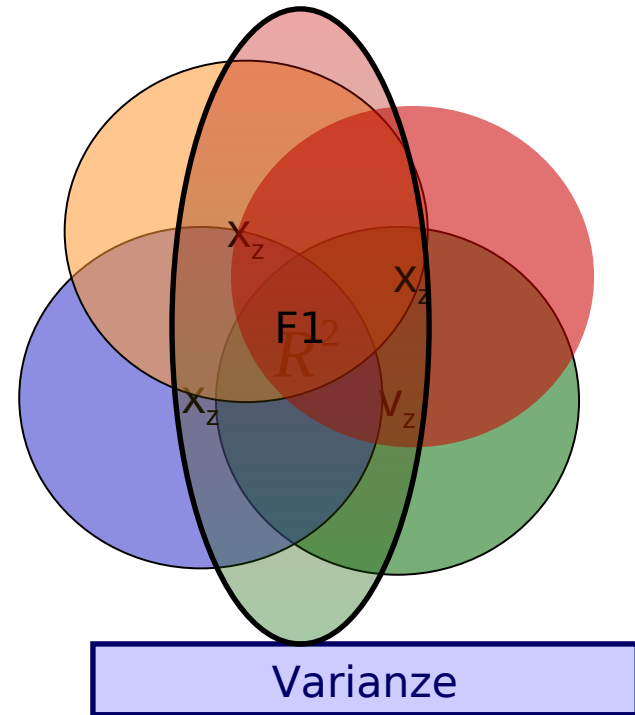
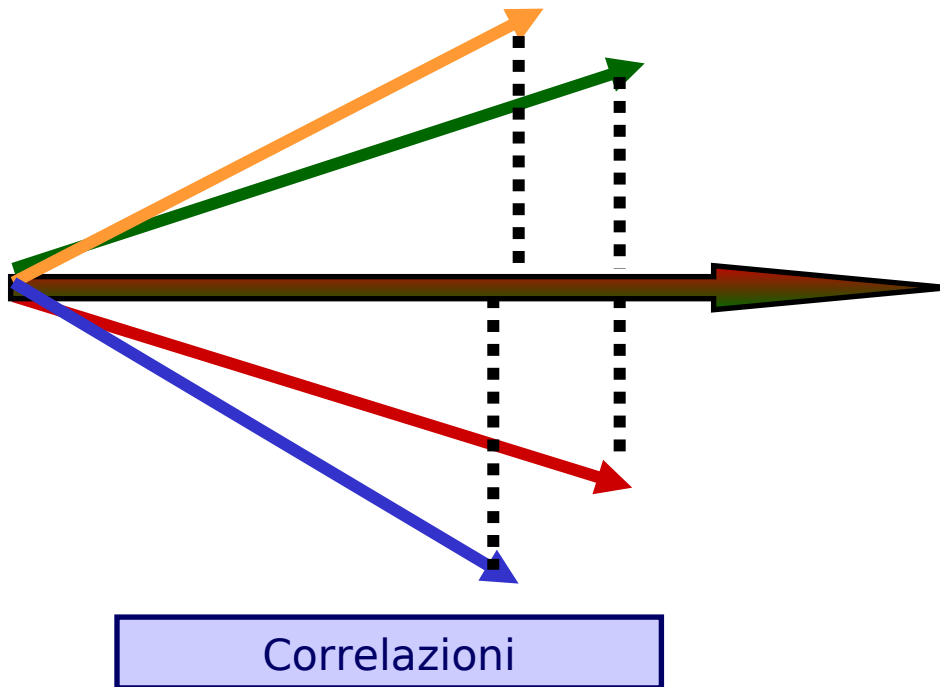
# Estrazione di più fattori

- Dunque saranno ortogonali



E le proiezioni saranno le correlazioni tra fattori e variabili (come per il caso di un fattore)

- La soluzione fattoriale si compone di:
  - La matrice di correlazione tra fattori e variabili
  - La varianza spiegata da ogni fattore



# Soluzione Fattoriale

- La soluzione fattoriale si compone di:
  - La matrice di correlazione tra fattori e variabili
  - La varianza spiegata da ogni fattore

Varianza totale spiegata

Componente	Autovalori iniziali			Pesi dei fattori non ruotati		
	Totale	% di varianza	% cumulata	Totale	% di varianza	% cumi
1	1.962	19.617	19.617	1.962	19.617	19.617
2	1.472	14.720	34.337	1.472	14.720	34.337
3	1.084	10.839	45.176			
4	.994	9.937	55.113			
5	.973	9.726	64.839			
6	.876	8.757	73.596			
7	.811	8.105	81.701			
8	.670	6.697	88.398			
9	.605	6.051	94.449			
10	.555	5.551	100.000			

Metodo di estrazione: Analisi componenti principali.

Matrice di componenti<sup>a</sup>


	Componente	
	1	2
v1	.438	.160
v2	.467	-.010
v3	.594	-.257
v4	.439	.610
v5	.453	.127
v6	.351	-.422
v7	.213	.454
v8	.304	-.397
v9	.722	-.210
v10	.135	.636

Metodo estrazione: analisi componenti principali

a. 2 componenti estratti

- La soluzione fattoriale si compone di:
  - La matrice di correlazione tra fattori e variabili

Le correlazioni tra fattori e variabili si  
chiamano  
PESI FATTORIALI o  
SATURAZIONI FATTORIALI

Matrice di componenti<sup>a</sup>

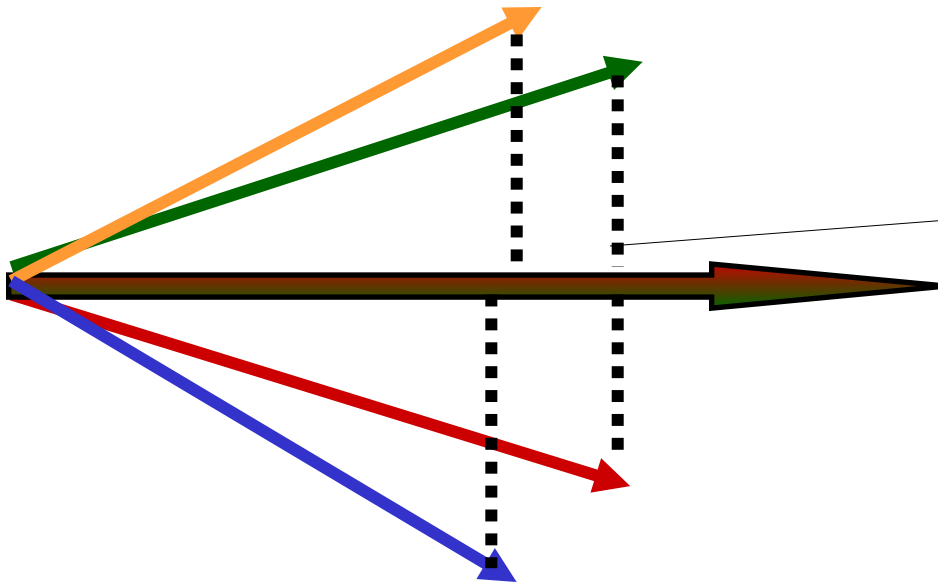
	Componente	
	1	2
v1	.438	.160
v2	.467	-.010
v3	.594	-.257
v4	.439	.610
v5	.453	.127
v6	.351	-.422
v7	.213	.454
v8	.304	-.397
v9	.722	-.210
v10	.135	.636

Metodo estrazione: analisi componenti prin

a. 2 componenti estratti

# Soluzione fattoriale

- La soluzione fattoriale rappresentata geometricamente può essere vista anche in una matrice numerica



Saturazioni = correlazioni Fattori  
variabili

	Componente	
	1	2
v1	.438	.160
v2	.467	-.010
v3	.594	-.257
v4	.439	.610
v5	.453	.127
v6	.351	-.422
v7	.213	.454
v8	.304	-.397
v9	.722	-.210
v10	.135	.636

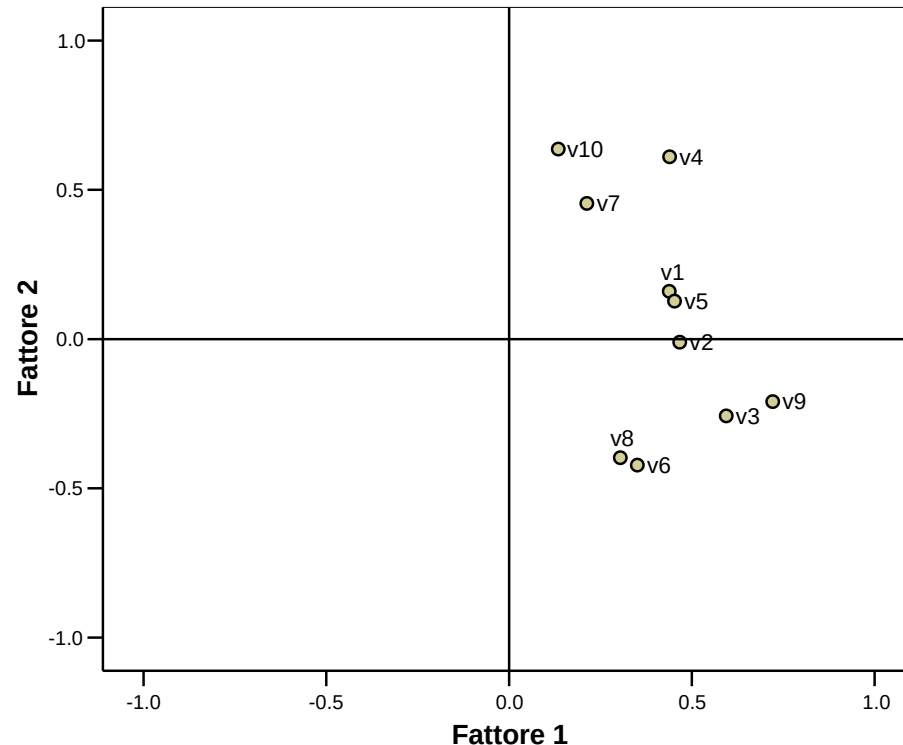
Metodo estrazione: analisi componenti principali.

a. 2 componenti estratti

# Plot dei Fattori

- Comunemente si visualizza lo spazio fattoriale mediante gli assi fattoriali e le variabili rappresentate come punti con coordinate uguali alle correlazioni con i fattori

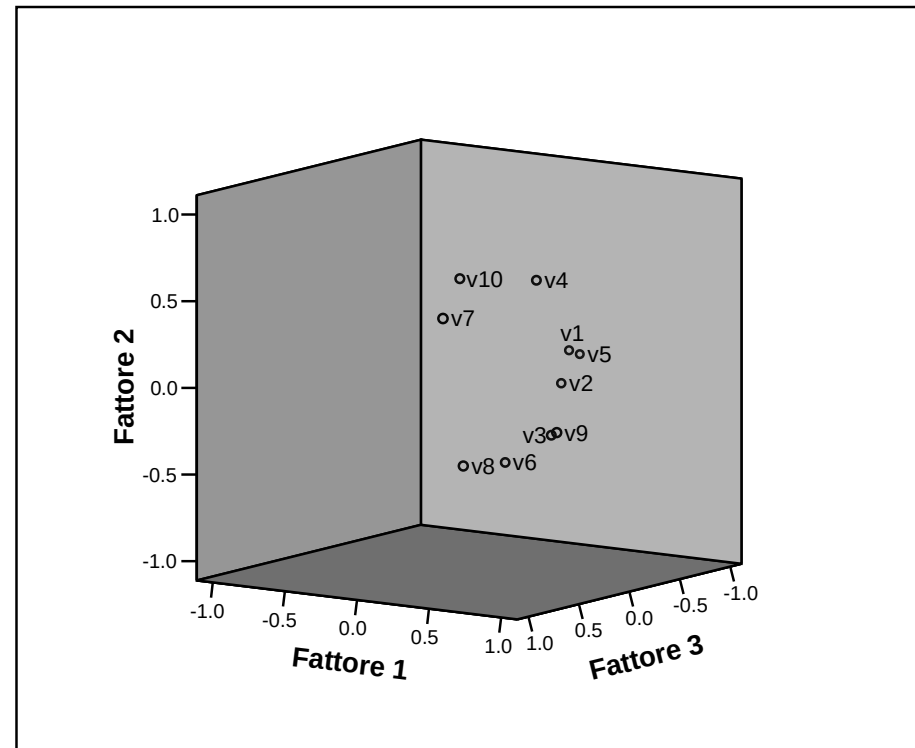
Grafico Fattori



# Plot K fattori

- L'estrazione di K fattori definisce uno spazio a K dimensioni dove tutte le variabili sono rappresentate mediante coordinate uguali alle correlazioni con i fattori

Grafico fattori



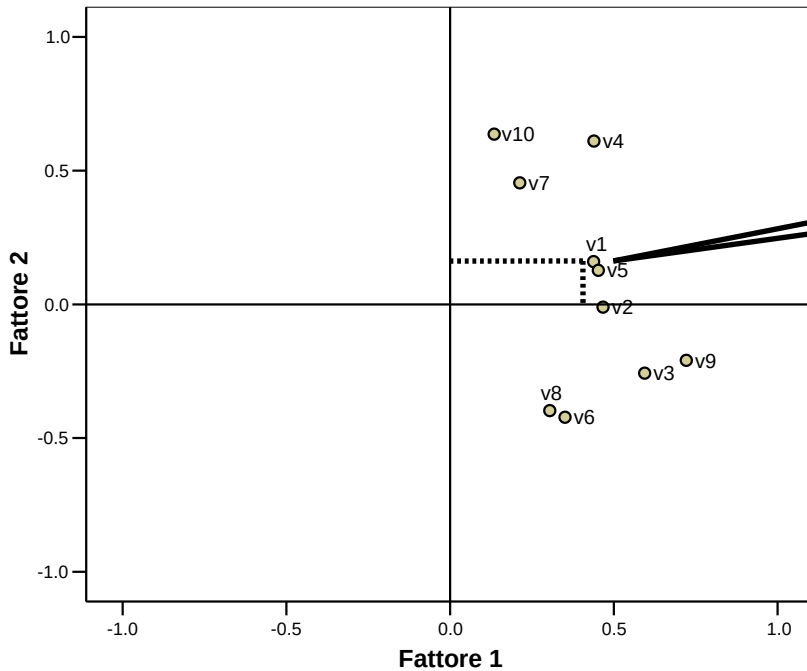
Per più di 3 fattori, si può visualizzare i fattori a 2 a 2



# Soluzione numerica

- La soluzione fattoriale rappresentata geometricamente può essere vista anche in una matrice numerica

Proiezioni = correlazioni Fattori  
variabili



Matrice di componenti<sup>a</sup>

	Componente	
	1	2
v1	.438	.160
v2	.467	-.010
v3	.594	-.257
v4	.439	.610
v5	.453	.127
v6	.351	-.422
v7	.213	.454
v8	.304	-.397
v9	.722	-.210
v10	.135	.636

Metodo estrazione: analisi componenti principali.

a. 2 componenti estratti

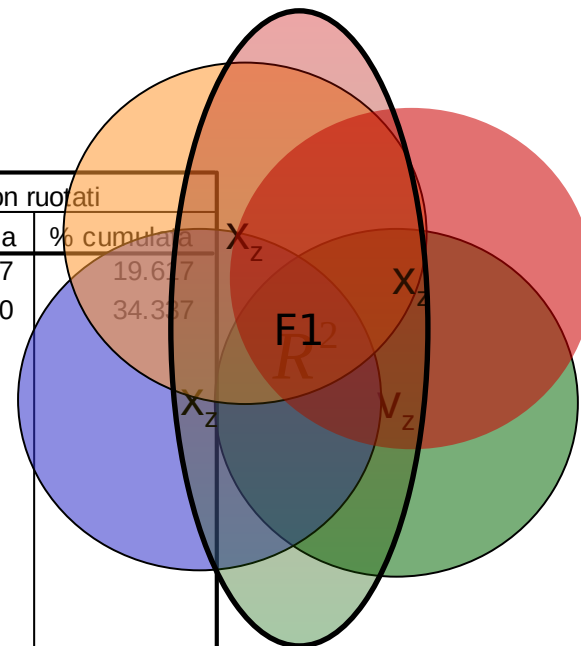
- La soluzione fattoriale si compone di:
  - La varianza spiegata da ogni fattore

Le varianze spiegate dai fattori si chiamano AUTOVALORI

Varianza totale spiegata

Componente	Autovalori iniziali			Pesi dei fattori non ruotati		
	Totale	% di varianza	% cumulata	Totale	% di varianza	% cumulata
1	1.962	19.617	19.617	1.962	19.617	19.617
2	1.472	14.720	34.337	1.472	14.720	34.337
3	1.084	10.839	45.176			
4	.994	9.937	55.113			
5	.973	9.726	64.839			
6	.876	8.757	73.596			
7	.811	8.105	81.701			
8	.670	6.697	88.398			
9	.605	6.051	94.449			
10	.555	5.551	100.000			

Metodo di estrazione: Analisi componenti principali.



Varianze

# Relazioni tra le informazioni

## SATURAZIONI FATTORIALI

Matrice di componenti<sup>a</sup>

	Componente	
	1	2
v1	.438	.160
v2	.467	-.010
v3	.594	-.257
v4	.439	.610
v5	.453	.127
v6	.351	-.422
v7	.213	.454
v8	.304	-.397
v9	.722	-.210
v10	.135	.636

Metodo estrazione: analisi componenti principali.

a. 2 componenti estratti

## Autovalori

Varianza totale spiegata

Componente	Autovalori iniziali			Pesi dei fattori no	
	Totale	% di varianza	% cumulata	Totale	% di varianz
1	1.962	19.617	19.617	1.962	19.61
2	1.472	14.720	34.337	1.472	14.72
3	1.084	10.839	45.176		
4	.994	9.937	55.113		
5	.973	9.726	64.839		
6	.876	8.757	73.596		
7	.811	8.105	81.701		
8	.670	6.697	88.398		
9	.605	6.051	94.449		
10	.555	5.551	100.000		

Metodo di estrazione: Analisi componenti principali.

La somma dei quadrati in colonna equivale alla varianza spiegata dal fattore corrispondente

## SATURAZIONI FATTORIALI

Matrice di componenti<sup>a</sup>

	Componente	
	1	2
v1	.438	.160
v2	.467	-.010
v3	.594	-.257
v4	.439	.610
v5	.453	.127
v6	.351	-.422
v7	.213	.454
v8	.304	-.397
v9	.722	-.210
v10	.135	.636

La somma dei quadrati in riga equivale alla varianza dell'item spiegata da tutti i fattori estratti

Tale quantità è denominata **comunalità**

Metodo estrazione: analisi componenti principali.

a. 2 componenti estratti

## SATURAZIONI FATTORIALI

Matrice di componenti<sup>a</sup>

	Componente		
	1	2	
v1	.438	.160	$.191 + .025$ .214
v2	.467	-.010	
v3	.594	-.257	
v4	.439	.610	
v5	.453	.127	
v6	.351	-.422	
v7	.213	.454	
v8	.304	-.397	
v9	.722	-.210	
v10	.135	.636	

I due fattori estratti spiegano il 21% della varianza dell'item

La comunalità indica quanto un item partecipa alla costruzione dei fattori

Metodo estrazione: analisi componenti principali.

a. 2 componenti estratti

# I passaggi fondamentali

- 1) Scelta del modello fattoriale
- 2) Scelta del numero di fattori
- 3) Rotazione della struttura fattoriale
- 4) Selezione delle variabili
- 5) Interpretazione dei fattori
- 6) Punteggi fattoriali

# Esempio (provvisorio)

La ricerca è volta a studiare le caratteristiche di personalità che possono associarsi a comportamenti pericolosi di adolescenti maschi. I costrutti di personalità misurati sono: il **sensation seeking**, la propensione al rischio, la mancanza di controllo. La tendenza ad attuare comportamenti pericolosi è misurata mediante un indice di frequenza di alcuni comportamenti indicatori (uso di droghe, partecipazione a gang, uso di armi, etc).

Il fine della ricerca è di stabilire se le variabili di personalità possono essere dei predittori della tendenza ad attuare comportamenti pericolosi.

## Descrizione dei dati

I tre costrutti di personalità sono stati misurati con 4 item ciascuno. Sensation seeking con gli item ss1, ss2, ss3 e ss4. Etc. etc. (da vedersi successivamente)

## Esempio (provvisorio)

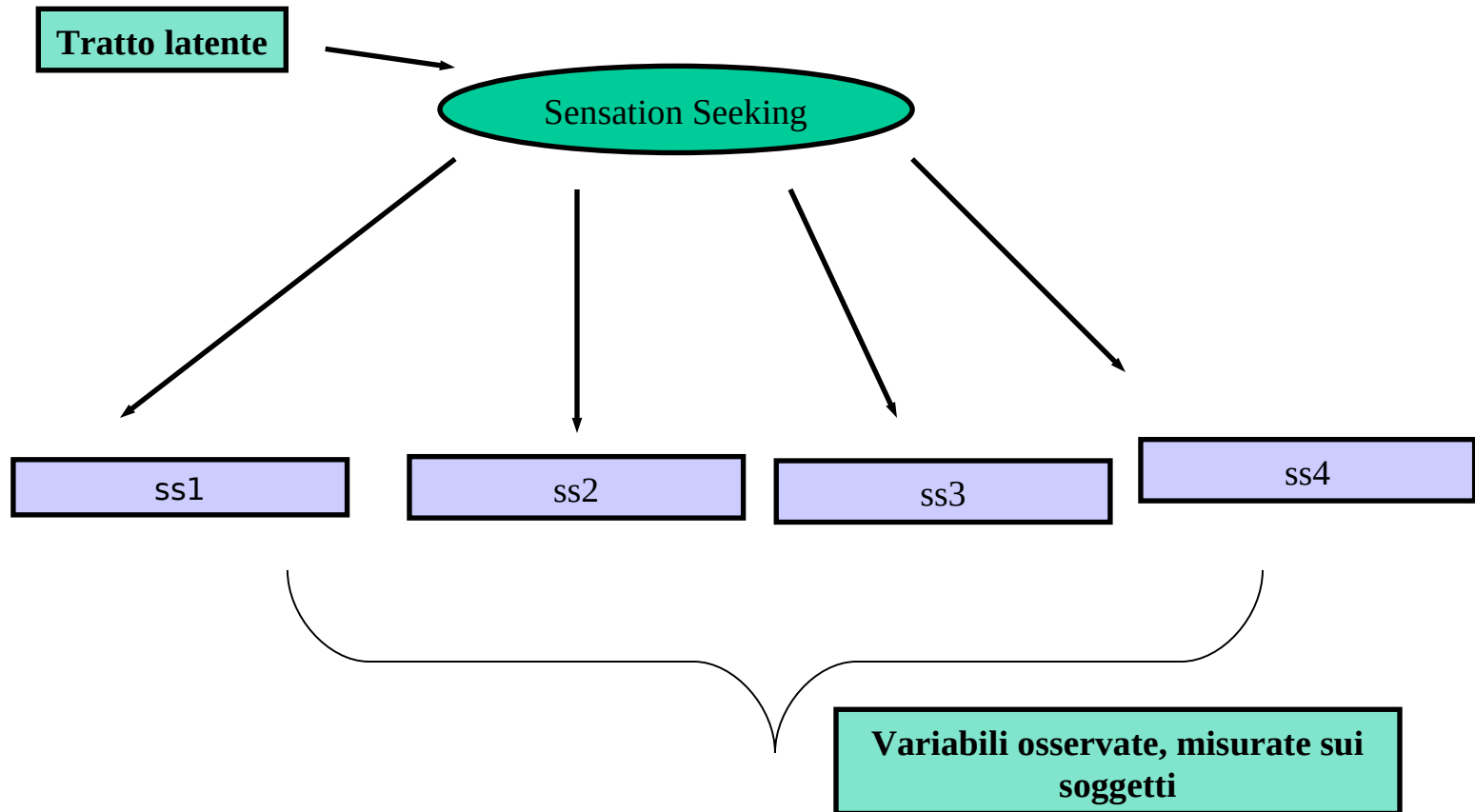
Intendiamo vedere se possiamo estrarre un fattore comune agli item di sensation seeking e vedere se tale fattore spiega bene la variabilità degli items



## Dunque (provvisorio)

- L'Analisi Fattoriale (e varianti) si propone di estrarre un numero limitato di fattori al fine di rappresentare al meglio la variabilità di tale set
- A tale scopo estrae una serie di fattori fra loro ortogonali al fine di massimizzare la correlazione fra variabili osservate
- La soluzione fattoriale è l'insieme di questi fattori, descritti dalle loro varianze (autovalori / numero item), le saturazioni fattoriali e la comunaltà degli item

# Un possibile modello



# Dunque (provvisorio)

Risposte standardizzate agli item (la standardizzazione è ininfluente)

	ss1	ss2	ss3	ss4
05	-.942	-1.027	.304	-1.366
	.013	.452	1.046	.490
	-.185	-.286	.550	.273
	-1.635	.318	-1.255	-1.813
	-1.087	.183	-.940	-1.437
06	.880	-.327	-1.524	-.362
08	1.672	.493	1.012	1.856
11	.724	-1.225	-.148	-1.754
14	-.097	1.061	.447	.854
18	.387	.510	1.041	.835
18	-.198	-.912	-.367	-.562
25	-1.487	-.764	.188	-1.178
25	-1.204	-.767	.184	-.667
13	-.584	.316	.277	1.286
16	-1.077	-.322	-.310	-.278
20	.537	.510	-.332	.651
28	1.099	.348	.793	-.786
21	-1.372	-.223	-.450	-1.545
30	.321	.405	.934	-.004
16	.549	.695	.568	.788
13	-.987	-1.455	-1.708	-2.047
11	1.218	1.901	2.011	.864
13	.566	-.841	-.848	.036
12	1.289	.398	.841	1.769
18	.244	.883	.400	.645
19	.822	1.290	1.137	.605

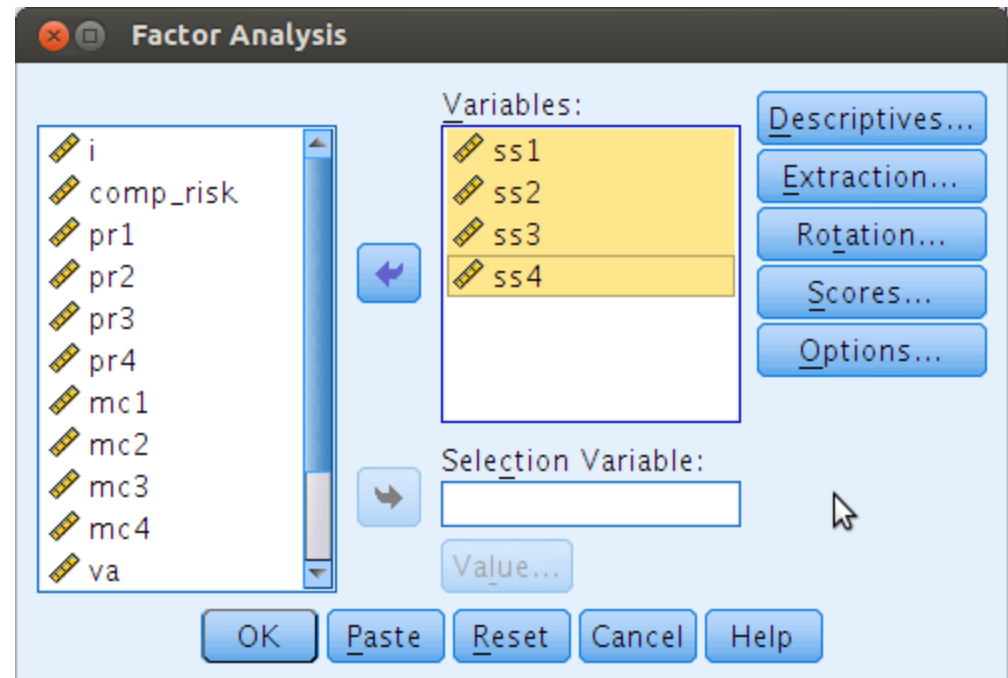
# SPSS

The screenshot shows the SPSS software interface. The 'Analyze' menu is open, and the 'Dimension Reduction' option is highlighted. A sub-menu is also open, showing 'Factor...', 'Correspondence Analysis...', and 'Optimal Scaling...'. The background shows a data table with columns 'comp\_risk', 'ss3', 'ss4', and 'pr1'.

	comp_risk	ss3	ss4	pr1
1	-3.0	.304	-1.366	-.419
2	1.0	1.046	.490	.303
3	1.0	.550	.273	.642
4	-3.0	-1.255	-1.813	-1.221
5	.0	-.940	-1.437	-.315
6	-1.0	-1.524	-.362	.655
7	1.0	1.012	1.856	1.454
8	.0	-.148	-1.754	-.443
9	3.0	.447	.854	1.467
10	3.0			.261
11	.0			-1.139
12	-.0			.703
13	1.0			.728
14	1.0			.431
15	-1.0			.772
16	-.0			.048
17	-7.0			-1.918
18	1.0			-.339
19	1.0			-.257
20	.0			.697
21	-.0			-1.007
22	.0			1.746
23	-.0			-1.662
24	-.222	1.289	.398	.841
25	1.388	.244	.883	.400
26	.119	.822	1.290	1.137
27	-2.350	-1.969	.736	-.363
28	1.658	.222	1.642	.486

# Dunque (provvisorio)

Seleziono le variabili che voglio analizzare



# Dunque (provvisorio)

Chiedo di estrarre un fattore  
(poi vedremo altri metodi per  
decidere quanti fattori estrarre)

**Factor Analysis: Extraction**

Method: **Principal components**

Analyze

- Correlation matrix
- Covariance matrix

Display

- Unrotated factor solution
- Screen plot

Extract

- Based on Eigenvalue  
Eigenvalues greater than:
- Fixed number of factors  
Factors to extract:

Maximum Iterations for Convergence:

**Continue** **Cancel** **Help**

Il primo fattore estratto (quello che ci interessa) spiega il 65% della varianza degli item

**Total Variance Explained**

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.615	65.382	65.382	2.615	65.382	65.382
2	.499	12.471	77.853			
3	.453	11.318	89.171			
4	.433	10.829	100.000			

Extraction Method: Principal Component Analysis.

# Soluzione fattoriale

Gli item sono ben correlati con il fattore, dunque possiamo utilizzare il fattore come variabile rappresentativa degli item

**Component  
Matrix<sup>a</sup>**

	Component
	1
ss1	.821
ss2	.798
ss3	.814
ss4	.800

Extraction Method:  
Principal  
Component  
Analysis.

a. 1  
component  
s extracted.



## Dunque (provvisorio)

- L'Analisi Fattoriale (e varianti) si propone di estrarre un numero limitato di fattori al fine di rappresentare al meglio la variabilità di tale set
- A tale scopo estrae una serie di fattori fra loro ortogonali al fine di massimizzare la correlazione fra variabili osservate
- La soluzione fattoriale è l'insieme di questi fattori, descritti dalle loro varianze (autovalori / numero item), le saturazioni fattoriali e la comunaltà degli item

Fine

