

Analisi Fattoriale

*Metodi Quantitativi per Economia,
Finanza e Management*

Esercitazione n°7

Metodi Quantitativi per Economia, Finanza e Management

Obiettivi di questa esercitazione:



Analisi Fattoriale

Tecnica di analisi multivariata

Quando si utilizza?

- Nel caso di un elevato numero di variabili quantitative, tra loro correlate (linearmente).
- **NB:** in contesti applicativi, è usata anche con variabili qualitative ordinali che esprimono scale di preferenza numeriche (scale di punteggi).

Perché si utilizza?

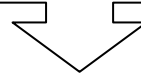
- Informazione condivisa tra le variabili correlate → è ridondante utilizzarle tutte
- Informazione dispersa tra le variabili → possibilità che le variabili, utilizzate singolarmente, siano poco esplicative



Analisi Fattoriale

OBIETTIVO

Sintetizzare le variabili originarie in un numero inferiore di variabili, dette fattori “LATENTI”



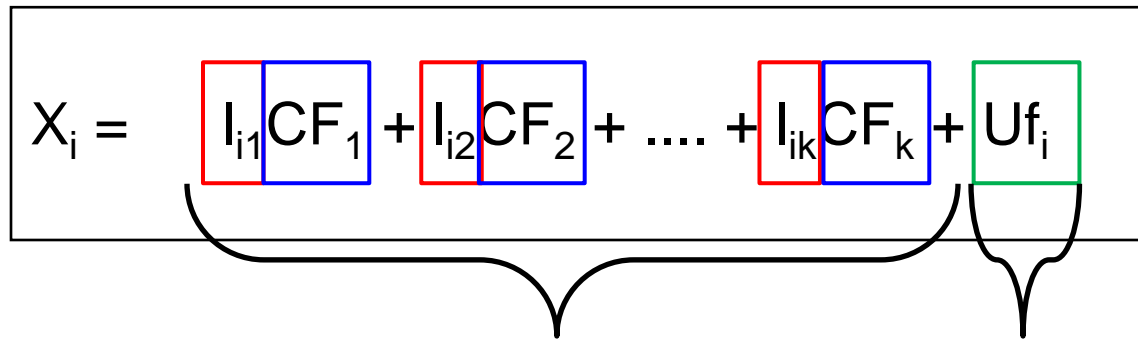
FATTORI LATENTI:

- concetti non direttamente misurabili
Esempio: la qualità della vita non è direttamente misurabile. Sono misurabili invece: il tasso di disoccupazione, tasso di aree verdi, tasso di inquinamento, aspettativa di vita...
- caratterizzati da una maggior facilità interpretativa
- spiegano «buona parte» della variabilità originaria, ovvero del contributo informativo delle variabili di partenza



Le ipotesi del Modello Fattoriale

Siano X_1, X_2, \dots, X_p variabili quantitative di partenza.
Ogni variabile X_i , con $i=1, \dots, p$, può essere espressa come:

$$X_i = I_{i1}CF_1 + I_{i2}CF_2 + \dots + I_{ik}CF_k + Uf_i$$


COMMON FACTORS

FACTOR LOADINGS

UNIQUE FACTOR

Info condivisa

Info specifica

$$\text{Var} [X_i] = \text{Communality} + \text{Var specifica}$$


porzione di varianza
spiegata complessivamente
dai fattori comuni



Metodo delle Componenti Principali

Una delle possibili tecniche per estrarre i fattori «latenti» (COMMON FACTORS) partendo dalle variabili originarie è il **metodo delle Componenti Principali**:

- tale metodo calcola i Common Factors come p nuove variabili, dette **Componenti Principali (CP)**, ottenute come combinazioni lineari delle variabili originali:

$$\boxed{CP_j = s_{j1}X_1 + s_{j2}X_2 + \dots + s_{jp}X_p} \quad \text{con } j=1, \dots, p$$

con $s_{j1}, s_{j2}, \dots, s_{jp}$ factor score coeff.

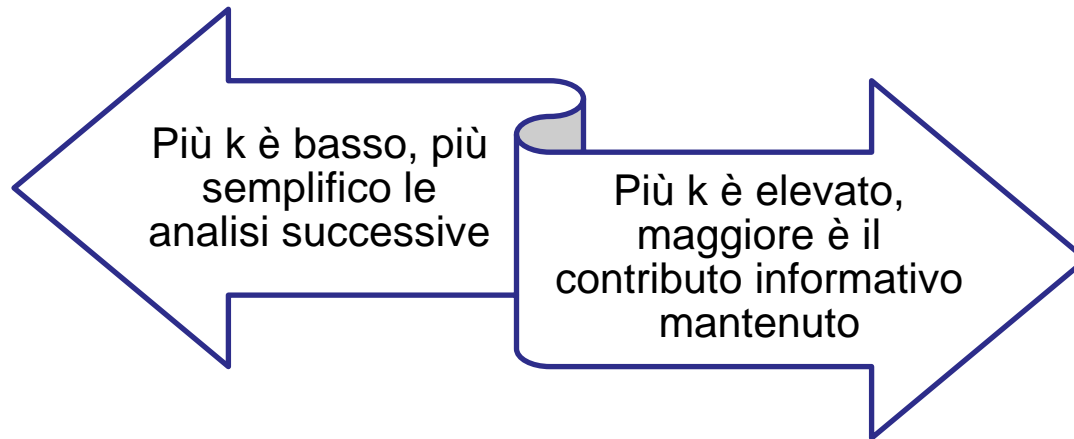
- Proprietà:
 - sono combinazioni lineari delle variabili di partenza
 - sono tra loro ortogonali (non correlate)
 - complessivamente spiegano la variabilità delle p variabili originarie
 - sono pari al numero delle variabili di partenza (p)
 - sono elencate in ordine decrescente rispetto alla variabilità spiegata



Metodo delle Componenti Principali

Se la correlazione tra le p variabili di partenza è elevata, un numero $k \ll p$ (k molto inferiore a p) di componenti principali è sufficiente a rappresentare in modo adeguato i dati originari, perché riassume una quota elevata della varianza totale.

Come determinare il numero k di fattori latenti tra le p componenti principali?



→ Per determinare il numero di fattori adeguato, è possibile ricorrere ad una serie di regole pratiche e strumenti grafici



Processo di analisi

Identificazione p variabili di partenza (variabili quantitative o scale di punteggio)



Selezione di k fattori
(dove $k < p$)

Utilizzo di alcuni criteri per la **selezione dei possibili valori di k** (è possibile identificare più valori di k adeguati)

Confronto tra le possibili soluzioni identificate (confronto delle comunalità)

Verifica dell'interpretabilità della soluzione scelta ed eventuale indagine di una soluzione differente



Interpretazione della soluzione finale



Metodi Quantitativi per Economia, Finanza e Management

Obiettivi di questa esercitazione:



Analisi Fattoriale: Esempio

Gli intervistati hanno espresso un giudizio sull'importanza di 21 caratteristiche relative a operatore/tariffa telefonica, utilizzando una scala da 1 a 10. (1=irrilevante, 10=fondamentale)

Immagine	1	2	3	4	5	6	7	8	9	10
Diffusione	1	2	3	4	5	6	7	8	9	10
Copertura della rete	1	2	3	4	5	6	7	8	9	10
.....										



Analisi Fattoriale: Esempio

VARIABILE	DESCRIZIONE
immagine_1	l'immagine dell'operatore
diffusione_1	la diffusione dell'operatore
copertura_1	la copertura della rete dell'operatore
assistenza_1	il servizio di assistenza dell'operatore
NoScattoRisp_1	l'assenza di scatto alla risposta
CostoSMS_1	il costo degli SMS
CostoMMS_1	il costo degli MMS
AccessoWeb_1	il costo di accesso a internet
NavigazioneWeb_1	il costo di navigazione in internet
ChiamateTuoOperatore_1	la possibilità di effettuare chiamate a costi inferiori verso numeri dello stesso operatore
SMSTuoOperatore_1	la possibilità inviare SMS a costi inferiori verso numeri dello stesso operatore
MMSTuoOperatore_1	la possibilità inviare MMS a costi inferiori verso numeri dello stesso operatore
vsPochiNumeri_1	le agevolazioni verso uno o più numeri di telefono
NumeriFissi_1	le agevolazioni verso numeri fissi
AltriOperatori_1	i costi verso altri operatori
Autoricarica_1	la possibilità di autoricarica
Promozioni_1	la possibilità di attivare promozioni sulle tariffe
ChiarezzaTariffe_1	la chiarezza espositiva delle tariffe
ComodatoUso_1	la possibilità di ricevere un cellulare in comodato d'uso
DurataMinContratto_1	la presenza di una durata minima del contratto
CambioTariffa_1	la facilità di cambiamento della tariffa



PRINCOMP – Sintassi generale (1/3)

Analisi fattoriale con il metodo delle componenti principali.

E' importante creare un subset contenente solo le variabili di interesse su cui applicare l'analisi delle componenti principali.

Per creare un nuovo dataset con le solo variabili di interesse, la sintassi è la seguente:

```
Nome_subset=nome_dataset[ ,c("var1",  
"var2", "var3", "varN", ...)]
```



PRINCOMP – Esempio (1/3)

Analisi fattoriale con il metodo delle componenti principali.

```
telefonia2=telefonia[,c("immagine_1", "diffusione_1",  
"copertura_1", "assistenza_1", "NoScattoRisp_1",  
"CostoSMS_1", "CostoMMS_1", "AccessoWeb_1",  
"NavigazioneWeb_1", "ChiamateTuoOperatore_1",  
"SMSTuoOperatore_1", "MMSTuoOperatore_1",  
"vsPochiNumeri_1", "NumeriFissi_1", "AltriOperatori_1",  
"Autoricarica_1", "Promozioni_1", "ChiarezzaTariffe_1",  
"ComodatoUso_1", "DurataMinContratto_1",  
"CambioTariffa_1")]
```



PRINCOMP – Sintassi generale (2/3)

Analisi fattoriale con il metodo delle componenti principali.

```
Nome_1 = princomp(nome_subset, cor=TRUE)
```

Opzione che indentifica che i calcoli devono esser svolti sulla matrice di correlazione



PRINCOMP – Esempio (2/3)

Analisi fattoriale con il metodo delle componenti principali.

```
fit <- princomp(telefonia2, cor=TRUE)
```

```
Summary(fit)
```



Per stampare l'output della funzione



Output PRINCOMP 1

```
> summary(fit)
```

```
Importance of components:
```

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	2.3487697	1.5538140	1.23058598	1.14118029	1.0278350
Proportion of Variance	0.2627009	0.1149685	0.07211152	0.06201393	0.0503069
Cumulative Proportion	0.2627009	0.3776694	0.44978090	0.51179482	0.5621017

	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
Standard deviation	0.99682506	0.97247004	0.95142066	0.94249143	0.87599373
Proportion of Variance	0.04731715	0.04503324	0.04310482	0.04229953	0.03654119
Cumulative Proportion	0.60941887	0.65445211	0.69755693	0.73985646	0.77639765

	Comp.11	Comp.12	Comp.13	Comp.14	Comp.15
Standard deviation	0.84099767	0.82212402	0.7662240	0.72173037	0.6883313
Proportion of Variance	0.03367986	0.03218514	0.0279571	0.02480451	0.0225619
Cumulative Proportion	0.81007751	0.84226265	0.8702198	0.89502427	0.9175862

	Comp.16	Comp.17	Comp.18	Comp.19	Comp.20
Standard deviation	0.66233460	0.63585205	0.58602435	0.54189851	0.439008161
Proportion of Variance	0.02088986	0.01925275	0.01635355	0.01398352	0.009177532
Cumulative Proportion	0.93847603	0.95772879	0.97408234	0.98806586	0.997243391

	Comp.21
Standard deviation	0.240600871
Proportion of Variance	0.002756609
Cumulative Proportion	1.000000000

21 variabili di partenza: X_1, X_2, \dots, X_{21} (immagine_1, diffusione_1,..)

La tecnica delle componenti principali determina in totale 21 componenti principali **CP₁, CP₂, ..., CP₂₁** tali che:

Somma varianza delle 21 componenti principali
=
Somma varianza delle 21 variabili originarie



Output PRINCOMP 1

```
> summary(fit)
```

```
Importance of components:
```

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	2.3487697	1.5538140	1.23058598	1.14118029	1.0278350
Proportion of Variance	0.2627009	0.1149685	0.07211152	0.06201393	0.0503069
Cumulative Proportion	0.2627009	0.3776694	0.44978090	0.51179482	0.5621017

	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
Standard deviation	0.99682506	0.97247004	0.95142066	0.94249143	0.87599373
Proportion of Variance	0.04731715	0.04503324	0.04310482	0.04229953	0.03654119
Cumulative Proportion	0.60941887	0.65445211	0.69755693	0.73985646	0.77639765

	Comp.11	Comp.12	Comp.13	Comp.14	Comp.15
Standard deviation	0.84099767	0.82212402	0.7662240	0.72173037	0.6883313
Proportion of Variance	0.03367986	0.03218514	0.0279571	0.02480451	0.0225619
Cumulative Proportion	0.81007751	0.84226265	0.8702198	0.89502427	0.9175862

	Comp.16	Comp.17	Comp.18	Comp.19	Comp.20
Standard deviation	0.66233460	0.63585205	0.58602435	0.54189851	0.439008161
Proportion of Variance	0.02088986	0.01925275	0.01635355	0.01398352	0.009177532
Cumulative Proportion	0.93847603	0.95772879	0.97408234	0.98806586	0.997243391

	Comp.21
Standard deviation	0.240600871
Proportion of Variance	0.002756609
Cumulative Proportion	1.000000000

In corrispondenza di ogni colonna/componente:

- **% PERCENTUALE di varianza spiegata dalla componente, sulla varianza totale**
- **% PERCENTUALE di VARIANZA CUMULATA (es: le prime 3 componenti spiegano il 45% della varianza totale)**



Quanti fattori considerare?

- **la regola autovalori > 1**
Selezione componenti principali con varianza maggiore di 1 (autovalori maggiori di 1) tenendo sotto controllo la % cumulata di varianza spiegata dalle componenti.
- **percentuale di varianza spiegata >60%**
- **lettura dello SCREE PLOT** (grafico di autovalore vs il numero di fattori)
Se il grafico mostra un “gomito” è plausibile ipotizzare l’esistenza di una struttura latente, se la forma è quasi rettilinea significa che i fattori sono solo una trasformazione delle variabili manifeste. I fattori rilevanti sono quelli al di sopra del gomito (a discrezione anche quello in corrispondenza del gomito).
- **rapporto tra numero di componenti e variabili**, il numero di fattori scelti dovrebbe essere circa 1/3 delle variabili originarie



PRINCOMP – Sintassi generale (3/3)

Analisi fattoriale con il metodo delle componenti principali.

Per calcolare gli **autovalori** dobbiamo scaricare un pacchetto e richiamarlo:

```
library("factoextra")
```

La funzione, invece, che calcola gli autovalori è:

```
Nome_2 = get_eigenvalue(nome_1)
```

dove *nome_1* è l'oggetto R che contiene la funzione *princomp*



PRINCOMP – Esempio (3/3)

Analisi fattoriale con il metodo delle componenti principali.

```
Eig.val <- get_eigenvalue(fit)
```

```
Eig.val
```



Per stampare l'output della funzione



Output PRINCOMP 1

```
> eig.val <- get_eigenvalue(fit)
> eig.val
```

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	5.51671913	26.2700911	26.27009
Dim.2	2.41433789	11.4968471	37.76694
Dim.3	1.51434184	7.2111516	44.97809
Dim.4	1.30229245	6.2013926	51.17948
Dim.5	1.05644480	5.0306895	56.21017
Dim.6	0.99366021	4.7317153	60.94189
Dim.7	0.94569797	4.5033237	65.44521
Dim.8	0.90520128	4.3104823	69.75569
Dim.9	0.88829009	4.2299528	73.98565
Dim.10	0.76736502	3.6541191	77.63977
Dim.11	0.70727708	3.3679861	81.00775
Dim.12	0.67588791	3.2185139	84.22627
Dim.13	0.58709920	2.7957105	87.02198
Dim.14	0.52089472	2.4804511	89.50243
Dim.15	0.47379998	2.2561904	91.75862
Dim.16	0.43868712	2.0889863	93.84760
Dim.17	0.40430783	1.9252754	95.77288
Dim.18	0.34342454	1.6353550	97.40823
Dim.19	0.29365400	1.3983524	98.80659
Dim.20	0.19272817	0.9177532	99.72434
Dim.21	0.05788878	0.2756609	100.00000

- Autovalore = VARIANZA della componente principale

Le numeriche della varianza percentuale e della varianza cumulata percentuale sono le stesse dell'output precedente (summary(fit))

Le 21 componenti principali $CP_1, CP_2, \dots, CP_{21}$



Regola autovalori >1

```
> eig.val <- get_eigenvalue(fit)
> eig.val
```

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	5.51671913	26.2700911	26.27009
Dim.2	2.41433789	11.4968471	37.76694
Dim.3	1.51434184	7.2111516	44.97809
Dim.4	1.30229245	6.2013926	51.17948
Dim.5	1.05644480	5.0306895	56.21017
Dim.6	0.99366021	4.7317153	60.94189
Dim.7	0.94569797	4.5033237	65.44521
Dim.8	0.90520128	4.3104823	69.75569
Dim.9	0.88829009	4.2299528	73.98565
Dim.10	0.76736502	3.6541191	77.63977
Dim.11	0.70727708	3.3679861	81.00775
Dim.12	0.67588791	3.2185139	84.22627
Dim.13	0.58709920	2.7957105	87.02198
Dim.14	0.52089472	2.4804511	89.50243
Dim.15	0.47379998	2.2561904	91.75862
Dim.16	0.43868712	2.0889863	93.84760
Dim.17	0.40430783	1.9252754	95.77288
Dim.18	0.34342454	1.6353550	97.40823
Dim.19	0.29365400	1.3983524	98.80659
Dim.20	0.19272817	0.9177532	99.72434
Dim.21	0.05788878	0.2756609	100.00000

Regola degli autovalori > 1
suggerisce di prendere in considerazione 5 fattori, che spiegano insieme il 56% della varianza totale.

%varianza
spiegata \approx 60%



Output PRINCOMP 2

Per ottenere il grafico dello Scree Plot, possiamo scrivere:

```
plot(nome_1, type='lines')
```

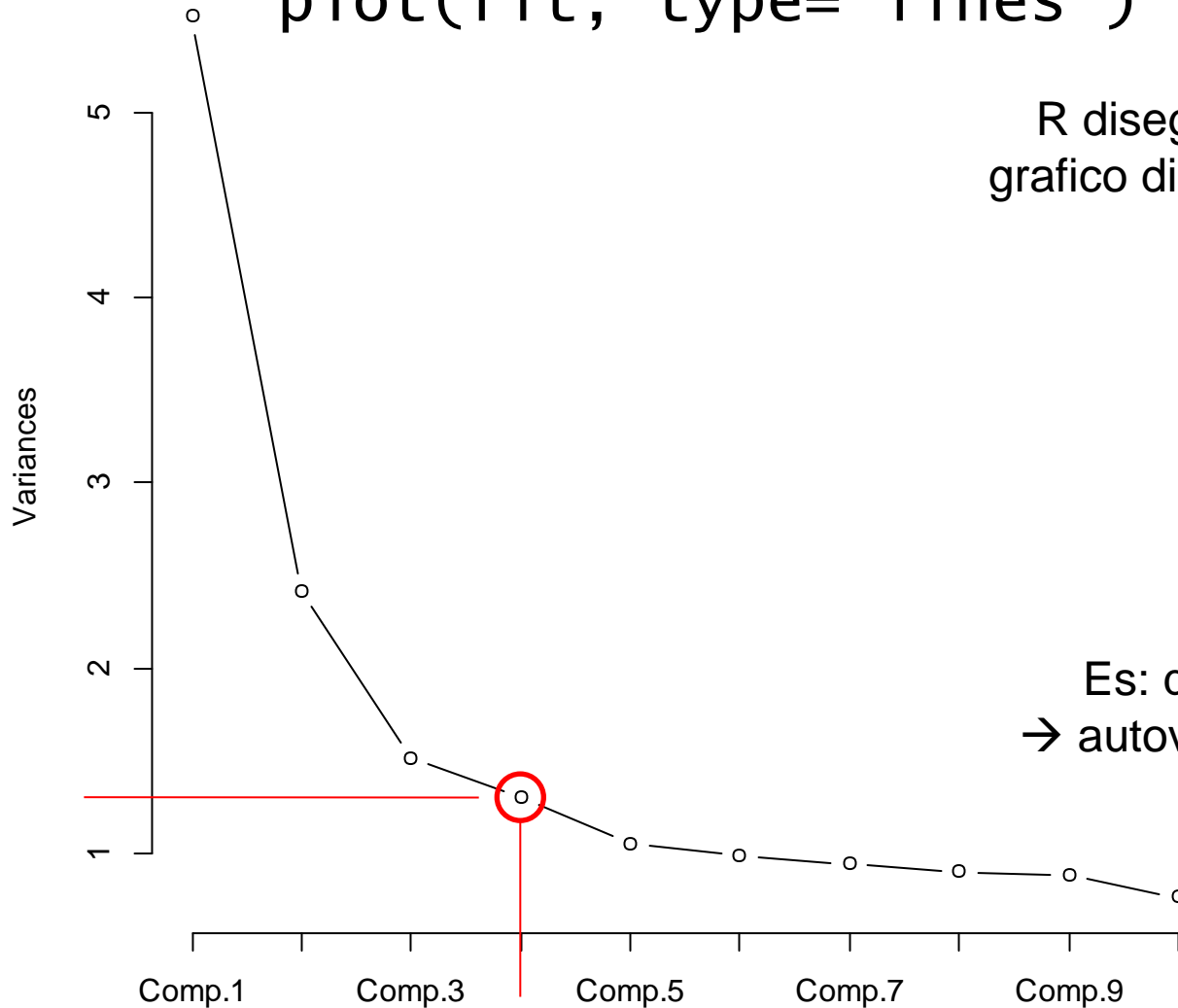
dove *nome_1* è l'oggetto R che contiene la funzione *princomp*



Output PRINCOMP 2

fit

```
plot(fit, type='lines')
```

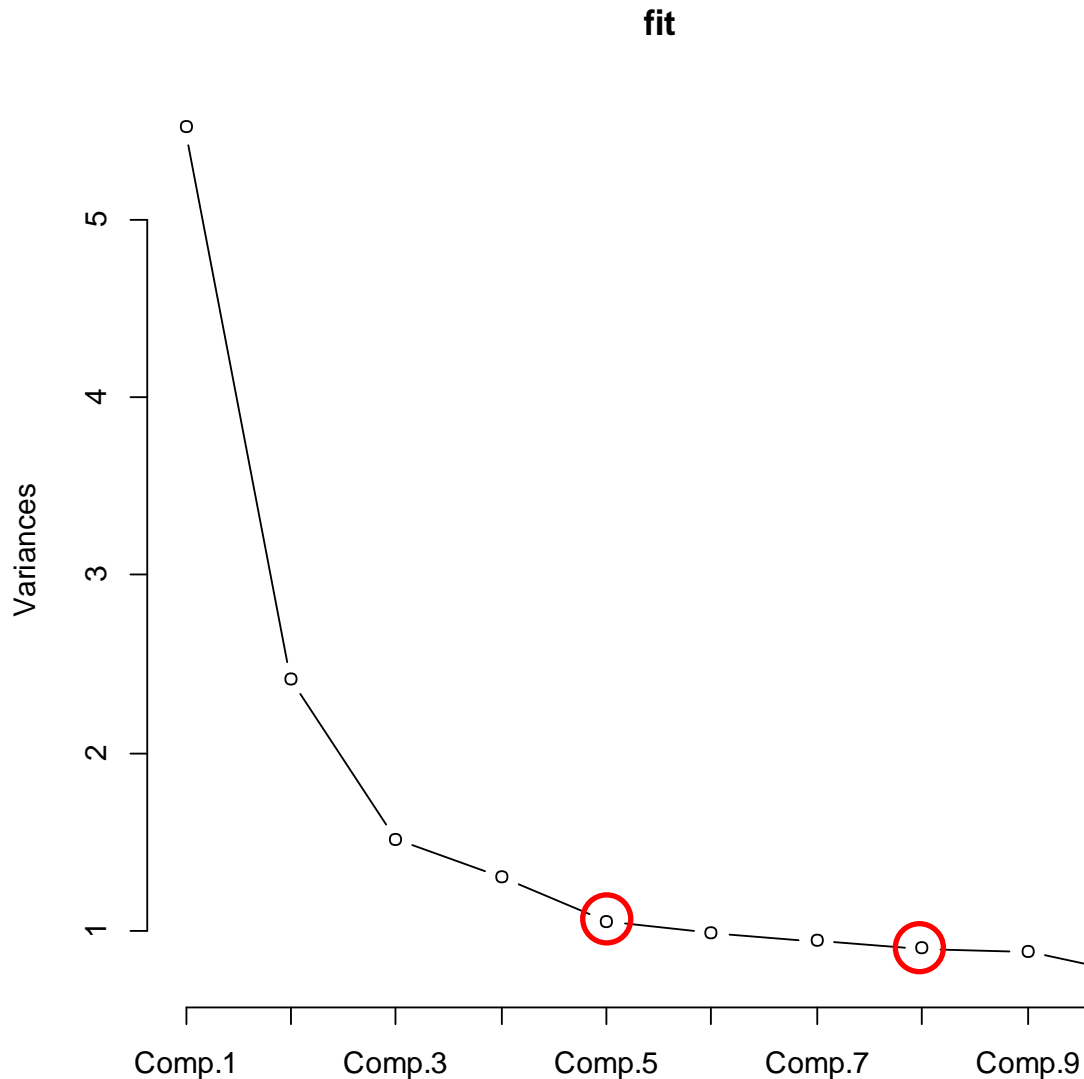


R disegna lo SCREE PLOT:
grafico di componente principale
vs autovalore

Es: quarta componente
→ autovalore=varianza=1.302



Lettura dello SCREE PLOT



REGOLA SCREE PLOT: Se il grafico mostra un “gomito” è plausibile ipotizzare l’esistenza di una struttura latente.

Lo scree plot mostra un gomito netto in corrispondenza di 5 fattori e uno in corrispondenza di 8 fattori.

% DI VARIANZA SPIEGATA:

- soluzione a 5 fattori: 56%
- soluzione a 8 fattori: 70%

**N°fattori = circa 1/3
variabili originali → circa
7 fattori**



principal - Sintassi

Quando abbiamo deciso un possibile numero di componenti da utilizzare, usiamo questa funzione per trovare i fattori latenti:

```
principal(nome_subset, residuals=FALSE,  
nfactors = num_fattori, rotate="none")
```

Nessuna rotazione

Nel nostro caso prendiamo il numero di autovalori >1, ovvero 5. Scriviamo:

```
z=principal(telefonía2, residuals=FALSE,  
nfactors = 5, rotate="none");
```

z -> stampa output

N.B. la funzione *principal* funziona richiamando la libreria *psych* (`library(psych)`)



principal - Output

```
> z
```

```
Principal Components Analysis
```

```
Call: principal(r = telefonia2, nfactors = 5, residuals = FALSE, rotate = "none")
```

```
Standardized loadings (pattern matrix) based upon correlation matrix
```

	PC1	PC2	PC3	PC4	PC5	h2	u2	com
immagine_1	0.33	-0.27	0.46	0.28	0.28	0.55	0.45	4.1
diffusione_1	0.31	0.08	0.70	0.20	0.33	0.75	0.25	2.1
copertura_1	0.18	0.36	0.43	0.41	-0.33	0.62	0.38	4.3
assistenza_1	0.44	0.31	0.33	-0.20	-0.42	0.62	0.38	4.2
NoScattoRisp_1	0.53	0.10	0.02	-0.10	0.26	0.37	0.63	1.6
CostoSMS_1	0.46	0.34	-0.15	-0.01	-0.05	0.35	0.65	2.1
CostoMMS_1	0.66	-0.35	-0.14	0.27	-0.21	0.69	0.31	2.3
AccessoWeb_1	0.58	-0.63	-0.16	0.11	0.03	0.78	0.22	2.2
NavigazioneWeb_1	0.58	-0.64	-0.13	0.03	0.08	0.77	0.23	2.1
ChiamateTuoOperatore_1	0.59	0.43	-0.20	0.34	0.05	0.69	0.31	2.8
SMSTuoOperatore_1	0.54	0.37	-0.26	0.31	-0.16	0.62	0.38	3.3
MMSTuoOperatore_1	0.66	-0.29	-0.11	0.29	-0.31	0.72	0.28	2.4
vsPochiNumeri_1	0.43	0.19	-0.15	0.29	0.43	0.51	0.49	3.4
NumeriFissi_1	0.51	0.02	-0.26	-0.18	0.23	0.42	0.58	2.2
AltriOperatori_1	0.60	0.37	-0.17	-0.23	-0.01	0.58	0.42	2.2
Autoricarica_1	0.58	0.02	0.13	-0.24	-0.01	0.41	0.59	1.5
Promozioni_1	0.58	0.20	0.03	-0.26	-0.03	0.45	0.55	1.7
ChiarezzaTariffe_1	0.49	0.25	0.00	-0.30	0.25	0.46	0.54	2.8
ComodatoUso_1	0.51	-0.36	0.09	-0.35	-0.18	0.55	0.45	3.0
DurataMinContratto_1	0.37	-0.42	0.36	-0.20	-0.09	0.49	0.51	3.5
CambioTariffa_1	0.53	0.25	0.18	-0.19	-0.01	0.41	0.59	2.0



principal – Output, loadings

Matrice dei FACTOR LOADINGS in corrispondenza della soluzione a 5 fattori

```
> z
Principal Components Analysis
Call: principal(r = telefonia2, nfactors = 5, residuals = FALSE, rotate = "none")
Standardized loadings (pattern matrix) based upon correlation matrix

  immagine_1  diffusione_1  copertura_1  assistenza_1  NoScattoRisp_1  CostoSMS_1  CostoMMS_1  AccessoWeb_1  NavigazioneWeb_1  ChiamateTuoOperatore_1  SMSTuoOperatore_1  MMSTuoOperatore_1  vsPochiNumeri_1  NumeriFissi_1  AltriOperatori_1  Autoricarica_1  Promozioni_1  ChiarezzaTariffe_1  ComodatoUso_1  DurataMinContratto_1  CambioTariffa_1
PC1  PC2  PC3  PC4  PC5  h2  u2  com
0.33 -0.27 0.46 0.28 0.28 0.55 0.45 4.1
0.31 0.08 0.70 0.20 0.33 0.75 0.25 2.1
0.18 0.36 0.43 0.41 -0.33 0.62 0.38 4.3
0.14 0.31 0.33 -0.20 -0.42 0.62 0.38 4.2
0.58 0.10 0.02 -0.10 0.26 0.37 0.63 1.6
0.46 0.34 -0.15 -0.01 -0.05 0.35 0.65 2.1
0.66 -0.35 -0.14 0.27 -0.21 0.69 0.31 2.3
0.58 -0.63 -0.16 0.11 0.03 0.78 0.22 2.2
0.58 -0.64 -0.13 0.03 0.08 0.77 0.23 2.1
0.59 0.43 -0.20 0.34 0.05 0.69 0.31 2.8
0.54 0.37 -0.26 0.31 -0.16 0.62 0.38 3.3
0.66 -0.29 -0.11 0.29 -0.31 0.72 0.28 2.4
0.43 0.19 -0.15 0.29 0.43 0.51 0.49 3.4
0.51 0.02 -0.26 -0.18 0.23 0.42 0.58 2.2
0.60 0.37 -0.17 -0.23 -0.01 0.58 0.42 2.2
0.58 0.02 0.13 -0.24 -0.01 0.41 0.59 1.5
0.58 0.10 0.03 -0.26 -0.03 0.45 0.55 1.7
0.49 0.26 0.00 -0.30 0.25 0.46 0.54 2.8
0.51 -0.36 0.09 -0.35 -0.18 0.55 0.45 3.0
0.37 -0.42 0.36 -0.20 -0.09 0.49 0.51 3.5
0.53 0.25 0.18 -0.19 -0.01 0.41 0.59 2.0
```

$$X_i = I_{i1}CF_1 + I_{i2}CF_2 + \dots + I_{ik}CF_k + U_{fi}$$

FACTOR LOADINGS

COMMON FACTOR

Es: Diffusione_1 = 0.31*CF1 + 0.08*CF2 + 0.70*CF3 +

Ciascun factor loading rappresenta la correlazione tra la variabile originaria e la componente principale

Es: Corr(Diffusione_1, CF1)=0.31



principal – Output, communality

z\$communality

```
> z$communality
```

immagine_1 0.5479127	diffusione_1 0.7452636	copertura_1 0.6225312
assistenza_1 0.6195847	NoScattoRisp_1 0.3680678	CostoSMS_1 0.3544767
CostoMMS_1 0.6861098	AccessoWeb_1 0.7792862	NavigazioneWeb_1 0.7688593
ChiamateTuoOperatore_1 0.6911749	SMSTuoOperatore_1 0.6207777	MMSTuoOperatore_1 0.7189244
vsPochiNumeri_1 0.5061591	NumeriFissi_1 0.4157727	AltriOperatori_1 0.5783326
Autoricarica_1 0.4127110	Promozioni_1 0.4508706	ChiarezzaTariffe_1 0.4572171
ComodatoUso_1 0.5547063	DurataMinContratto_1 0.4934322	CambioTariffa_1 0.4119655

$$\text{Var}[X_i] = \text{Communality} + \text{Var specifica}$$

Comunalità della variabile
Immagine_1 = 0.547
(Porzione della varianza della
variabile Immagine_1 spiegata dai 5
fattori scelti)

porzione di varianza
spiegata complessivamente
dai fattori comuni



Metodi Quantitativi per Economia, Finanza e Management

Obiettivi di questa esercitazione:



Confronto Soluzioni

Confrontiamo la soluzione a 5 e a 8 fattori.

```
principal(telefonía2, residuals=FALSE  
nfactors = 8, rotate="none")
```



Confronto Comunalità

Variabile	COMUNALITA' FINALI	
	n=5	n=8
immagine_1	0.55	0.69
diffusione_1	0.75	0.79
copertura_1	0.62	0.73
assistenza_1	0.62	0.71
NoScattoRisp_1	0.37	0.59
CostoSMS_1	0.35	0.70
CostoMMS_1	0.69	0.79
AccessoWeb_1	0.78	0.83
NavigazioneWeb_1	0.77	0.82
ChiamateTuoOperatore_1	0.69	0.75
SMSTuoOperatore_1	0.62	0.74
MMSTuoOperatore_1	0.72	0.82
vsPochiNumeri_1	0.51	0.80
NumeriFissi_1	0.42	0.54
AltriOperatori_1	0.58	0.65
Autoricarica_1	0.41	0.62
Promozioni_1	0.45	0.59
ChiarezzaTariffe_1	0.46	0.60
ComodatoUso_1	0.55	0.68
DurataMinContratto_1	0.49	0.70
CambioTariffa_1	0.41	0.51
Totale	11.80	14.65

Analisi della varianza spiegata dai fattori (comunalità finali)

Per ogni variabile si evidenziano le celle in corrispondenza delle quali la comunalità aumenta in maniera sostanziale per effetto dell'estrazione di ulteriori fattori (dalla soluzione a 5 fattori alla soluzione a 8 fattori).

Occorre verificare anche che in corrispondenza della soluzione da preferire tutte le variabili risultino sufficientemente «spiegate».

→ Scegliamo quella a 8 fattori



Metodi Quantitativi per Economia, Finanza e Management

Obiettivi di questa esercitazione:



Interpretazione Fattori

Una volta estratti, i fattori vanno interpretati.

Una rotazione ortogonale nello spazio dei fattori non influenza la validità del modello: sfruttiamo questa caratteristica per ottenere dei fattori più facilmente interpretabili!

Dobbiamo fare in modo che ognuna delle variabili originali sia molto correlata con al massimo un fattore e poco correlata con gli altri.



Metodi di rotazione

La rotazione opera sulla matrice dei loadings.

Esistono diversi metodi, tra cui:

1. METODO VARIMAX: minimizza il numero di variabili che hanno correlazioni alte con un fattore
2. METODO QUARTIMAX: minimizza il numero di fattori che hanno correlazioni alte con una variabile
3. METODO EQUIMAX: è una combinazione dei due metodi precedenti

IMPORTANTE: la % di varianza originaria, spiegata complessivamente dei fattori ruotati, rimane inalterata, mentre si modifica la % di varianza spiegata da ciascun fattore



Principal - Rotazione

Operiamo una rotazione dei fattori con il metodo Varimax.

```
principal(nome_subset, residuals=FALSE,  
nfactors = num_fattori, rotate = "varimax")
```

Rotazione varimax

Nel nostro caso scriviamo:

```
z3=principal(telefonია2, residuals=FALSE, nfactors = 8, rotation='varimax')
```

```
print(z3$loadings, sort=TRUE, cutoff=0.3)
```

Rotazione opera
sulle loadings

Stampa le loadings sopra il
valore indicato dal *cutoff*

Visualizzare l'output
della rotazione



Principal – Rotazione - Output

Operiamo una rotazione dei fattori con il metodo Varimax.

```
> z3=principal(telefonia2, nfactors = 8, residuals = FALSE, rotate="varimax")
> print(z3$loadings, sort=TRUE, cutoff=0.3)
```

Loadings:

	RC2	RC6	RC1	RC8	RC7	RC4	RC3	RC5
CostoMMS_1	0.815							
AccessoWeb_1	0.774		0.305					
NavigazioneWeb_1	0.722		0.331					
MMSTuoOperatore_1	0.810							
Autoricarica_1	0.304	0.638						
Promozioni_1		0.684						
ChiarezzaTariffe_1		0.692						
NoScattoRisp_1			0.664					
NumeriFissi_1			0.547					0.349
AltriOperatori_1			0.622	0.352				
CostoSMS_1				0.780				
SMSTuoOperatore_1				0.770				
ComodatoUso_1					0.738			
DurataMinContratto_1					0.780			
copertura_1						0.833		
assistenza_1			0.355			0.523		-0.401
immagine_1							0.758	
diffusione_1		0.312					0.780	
vsPochiNumeri_1								0.842
ChiamateTuoOperatore_1			0.489	0.486		0.331		0.325
CambioTariffa_1			0.325		0.404			
SS loadings	2.864	2.028	2.001	1.865	1.704	1.527	1.447	1.212
Proportion Var	0.136	0.097	0.095	0.089	0.081	0.073	0.069	0.058



Analisi Fattoriale: Esempio

VARIABILE	DESCRIZIONE
immagine_1	l'immagine dell'operatore
diffusione_1	la diffusione dell'operatore
copertura_1	la copertura della rete dell'operatore
assistenza_1	il servizio di assistenza dell'operatore
NoScattoRisp_1	l'assenza di scatto alla risposta
CostoSMS_1	il costo degli SMS
CostoMMS_1	il costo degli MMS
AccessoWeb_1	il costo di accesso a internet
NavigazioneWeb_1	il costo di navigazione in internet
ChiamateTuoOperatore_1	la possibilità di effettuare chiamate a costi inferiori verso numeri dello stesso operatore
SMSTuoOperatore_1	la possibilità inviare SMS a costi inferiori verso numeri dello stesso operatore
MMSTuoOperatore_1	la possibilità inviare MMS a costi inferiori verso numeri dello stesso operatore
vsPochiNumeri_1	le agevolazioni verso uno o più numeri di telefono
NumeriFissi_1	le agevolazioni verso numeri fissi
AltriOperatori_1	i costi verso altri operatori
Autoricarica_1	la possibilità di autoricarica
Promozioni_1	la possibilità di attivare promozioni sulle tariffe
ChiarezzaTariffe_1	la chiarezza espositiva delle tariffe
ComodatoUso_1	la possibilità di ricevere un cellulare in comodato d'uso
DurataMinContratto_1	la presenza di una durata minima del contratto
CambioTariffa_1	la facilità di cambiamento della tariffa



fa.diagram – Rotazione - Output

Per visualizzare in modo automatico i vari raggruppamenti dei fattori latenti possiamo usare questa funzione:

```
fa.diagram(nome_oggetto_analisi_fattoriale)
```

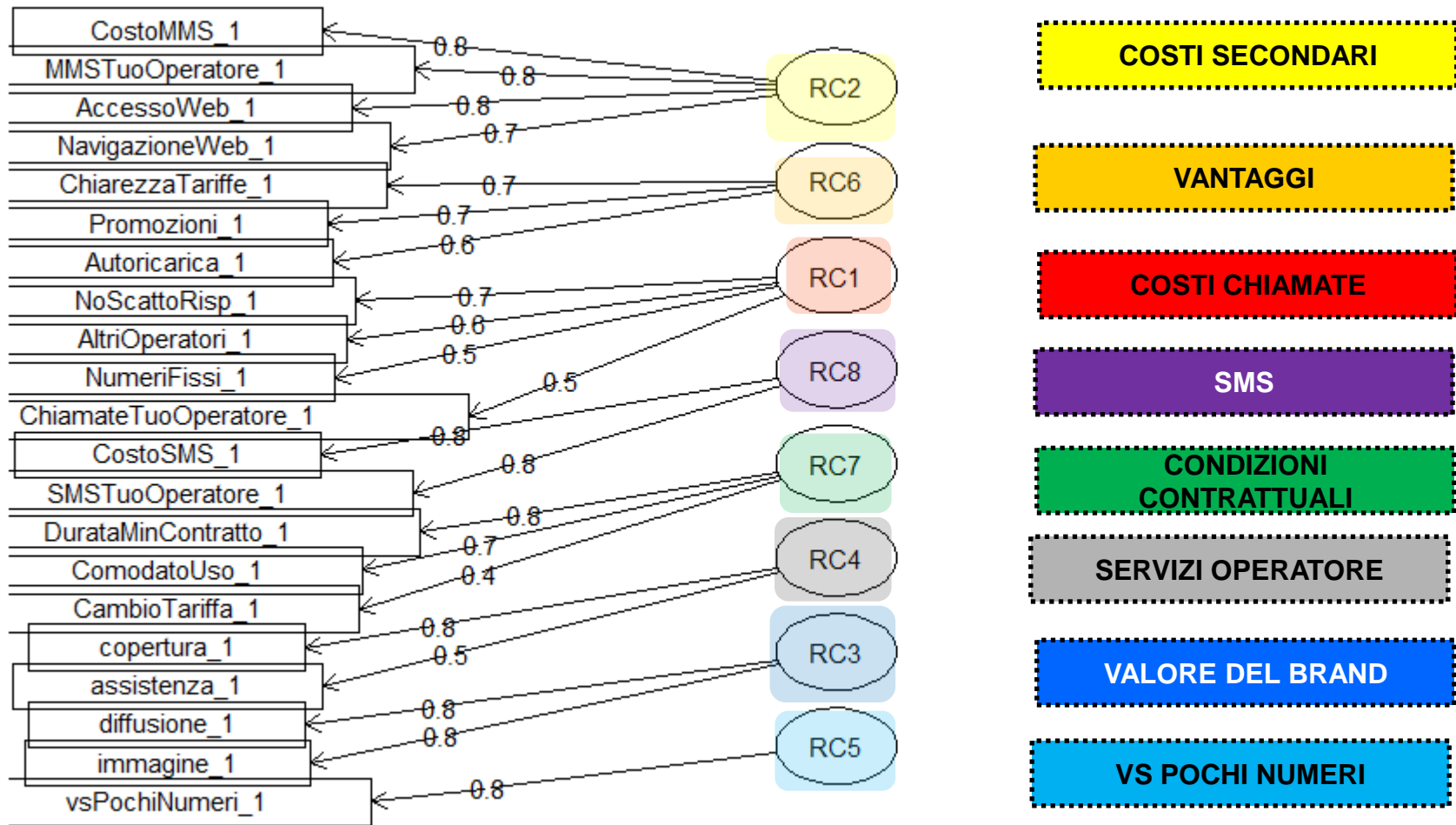
Nel nostro caso scriviamo:

```
fa.diagram(z3)
```



fa.diagram – Rotazione - Output

Factor Analysis



Salvare i fattori latenti

Per salvare i nuovi fattori latenti trovati dall'analisi fattoriale è necessario aggiungere un'opzione alla funzione **principal** e poi aggiungere le nuove variabili/colonne nel dataset principale, creandone quindi uno nuovo.

```
New_name=principal(nome_subset, residuals=FALSE,  
nfactors = num_fattori, rotate = 'varimax',  
score=TRUE)
```

Salva gli scores per ogni osservazione del dataset
sugli n fattori latenti individuati

Nuovo dataset con
le nuove colonne
contenenti gli
scores

```
New_name_2=cbind(dataset_principale, new_name$scores)
```



Salvare i fattori latenti - Esempio

```
z3=principal(telefonía2, nfactórs = 8, residuals = FALSE, rotate="varimax", score=TRUE)
```

```
telefonía_scored=cbind(telefonía2, z3$score)
```

```
> z3=principal(telefonía2, nfactórs = 8, residuals = FALSE, rotate="varimax", score=TRUE)
> telefonía_scored=cbind(telefonía2,z3$score)
> names(telefonía_scored)
 [1] "immagine_1"          "diffusione_1"          "copertura_1"          "assistenza_1"
 [5] "NoScattoRisp_1"      "CostoSMS_1"           "CostoMMS_1"          "AccessoWeb_1"
 [9] "NavigazioneWeb_1"    "ChiamateTuoOperatore_1" "SMSTuoOperatore_1"    "MMSTuoOperatore_1"
[13] "vsPochiNumeri_1"     "NumeriFissi_1"        "AltriOperatori_1"    "Autoricarica_1"
[17] "Promozioni_1"        "ChiarezzaTariffe_1"   "ComodatoUso_1"       "DurataMinContratto_1"
[21] "CambioTariffa_1"     "RC2"                  "RC6"                 "RC1"
[25] "RC8"                 "RC7"                  "RC4"                 "RC3"
[29] "RC5"
> fix(telefonía_scored)
```



Fattori

- Una volta scelta la soluzione ottimale, è possibile utilizzare i fattori ottenuti come nuove “macro-variabili” da inserire in ulteriori analisi sul fenomeno indagato, al posto delle variabili originarie;
- Nel file di dati si potranno aggiungere 8 nuove variabili:
 - **Costi secondari,**
 - **Vantaggi,**
 - **Costi chiamate,**
 - **SMS,**
 - **Condizioni contrattuali,**
 - **Servizi Operatore,**
 - **Valore del Brand,**
 - **Vs pochi numeri.**



Come rinominare i fattori

- Scaricare e richiamare il pacchetto plyr

```
library(plyr)
```

```
telefonia_scored=rename(telefonia_scored,  
c("RC2"="costi_secondari",  
"RC6"="vantaggi",...))
```



Riepilogo del processo

Step di analisi (1/2)

STEP 1: scegliere quanti fattori considerare (scelta di varie soluzioni)

- la regola autovalori > 1
- lettura dello SCREE PLOT
- Circa 1/3 delle variabili originarie
- Variabilità spiegata $> 60\%$

```
Nome_1 = princomp(nome_subset, cor=TRUE)
```

```
get_eigenvalue(nome_1)
```

```
plot(nome_1, type='lines')
```

STEP 2: confrontare le soluzioni scelte

- cumunalità finali

```
Nome_2= principal(nome_subset, residuals=FALSE, nfactors = num_fattori,  
rotate = 'none')
```

```
Nome_2$communality
```



Step di analisi (2/2)

STEP 3: una volta scelta la soluzione finale

- ruotare i fattori
- interpretare i fattori
- salvare il data set con i fattori

```
nome_3= principal(nome_subset, residuals=FALSE, nfactors = num_fattori,  
rotate = 'varimax', score=TRUE)
```

```
print(nome_3$loadings, sort=TRUE, cutoff=0.3)
```

```
fa.diagram(nome_3)
```

```
Nome_4=cbind(dataset_originale, nome_2$scores)
```

STEP 4: se l'interpretazione non è soddisfacente ripetere lo step n°3 variando metodo di rotazione o provando un'altra soluzione.

