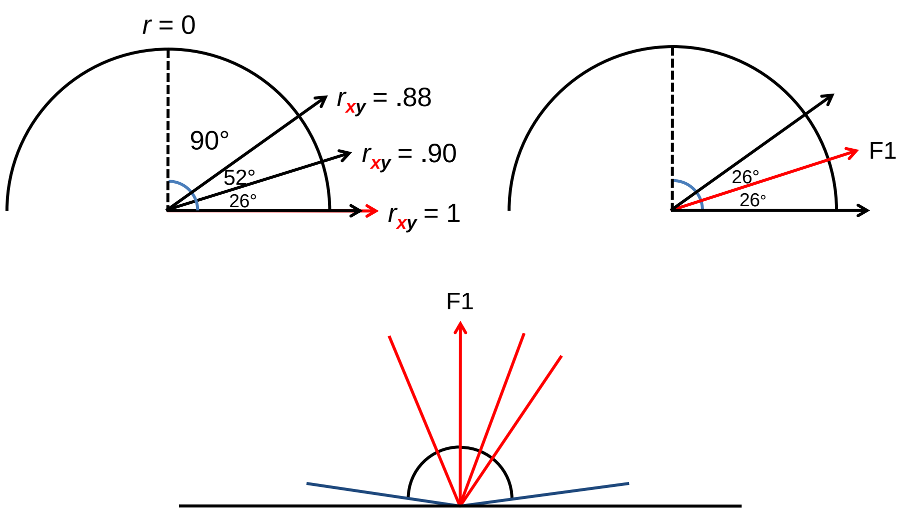
METODI E TECNICHE DI ANALISI DEI DATI II
E
LABORATORIO

AA 2016/2017

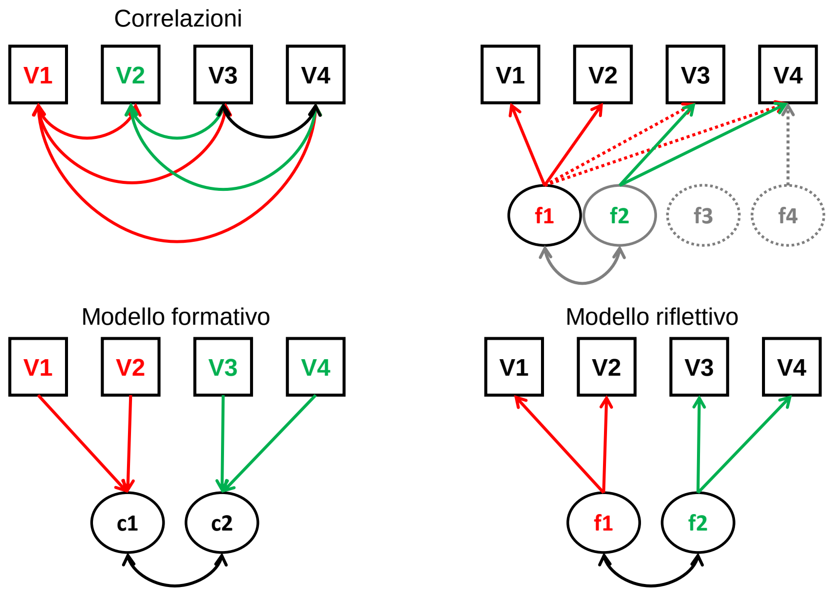
PROF. V.P. SENESE

ANALISI FATTORIALE



Dancey & Reidy, 2007

ANALISI FATTORIALE



ANALISI FATTORIALE

$$r_{xy} = \frac{\text{cov}_{xy}}{ds_x \times ds_y}$$

var-cov

	x1	x2	x3	x4	x5	x6
x1	0.81	0.81	0.82	0.67	1.14	0.39
x2	0.81	1.72	0.77	0.50	1.42	0.16
x3	0.82	0.77	1.91	1.03	1.60	0.62
x4	0.67	0.50	1.03	9.39	10.10	8.27
x5	1.14	1.42	1.60	10.10	23.13	8.06
x6	0.39	0.16	0.62	8.27	8.06	15.54

$$r_{xy} = \frac{.81}{.9 \times 1.3115} = .68 \longrightarrow$$

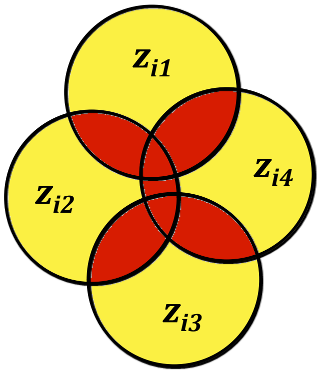
$$ds_x = \sqrt{.81} = .9$$

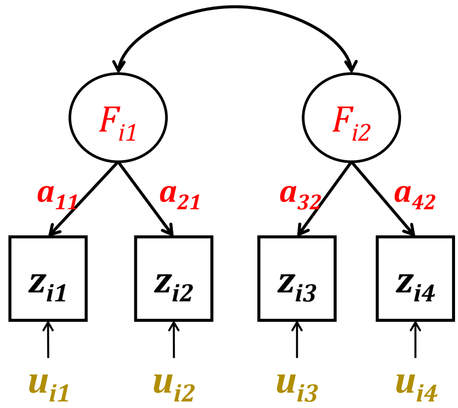
$$ds_y = \sqrt{1.72} = 1.3115$$

corr

	x1	x2	x3	x4	x5	x6
x1	1.00	0.68	0.66	0.24	0.26	0.11
x2	0.68	1.00	0.43	0.13	0.23	0.03
x3	0.66	0.43	1.00	0.24	0.24	0.11
x4	0.24	0.13	0.24	1.00	0.69	0.68
x5	0.26	0.23	0.24	0.69	1.00	0.42
x6	0.11	0.03	0.11	0.68	0.42	1.00

ANALISI FATTORIALE





$$z_{ik} = F_{i1}a_{k1} + F_{i2}a_{k2} + \ldots + F_{im}a_{km} + u_{ik}$$

$\boldsymbol{F_{i1}}$ = punteggio **fattoriale**
$\boldsymbol{a_{k1}}$ = saturazione
$\boldsymbol{u_{ik}}$ = punteggio **unico** o **fattore unico**

ANALISI FATTORIALE

$\boldsymbol{n}$ = soggetti
$\boldsymbol{m}$ = fattori
$\boldsymbol{k}$ = variabili

$$z_{ik} = F_{i1}a_{k1} + F_{i2}a_{k2} + \ldots + F_{im}a_{km} + u_{ik}$$

$$\begin{vmatrix} z_{11} & z_{12} & z_{1k} \\ z_{21} & z_{21} & z_{2k} \\ z_{31} & z_{31} & z_{nk} \end{vmatrix} = \begin{vmatrix} F_{11} & F_{12} & F_{1m} \\ F_{21} & F_{21} & F_{2m} \\ F_{31} & F_{31} & F_{nm} \end{vmatrix} \cdot \begin{vmatrix} a_{11} & a_{21} & a_{k1} \\ a_{12} & a_{22} & a_{k2} \\ a_{13} & a_{23} & a_{km} \end{vmatrix} + \begin{vmatrix} u_{11} & u_{12} & u_{1k} \\ u_{21} & u_{21} & u_{2k} \\ u_{31} & u_{31} & u_{nk} \end{vmatrix}$$

$$Z = FA' + U$$

$\boldsymbol{F_{i1}}$ = punteggio fattoriale
$\boldsymbol{a_{k1}}$ = saturazione
$\boldsymbol{u_{ik}}$ = **unicità**

ANALISI FATTORIALE

$$Var.Totale = Var.Comune + Var.Unica$$

$$Var.Totale = \underbrace{h^2}_{\text{Comunalità}} + \underbrace{u^2}_{\text{Unicità}}$$

ASSUNTI ANALISI FATTORIALE

- I **fattori unici** non correlano con i fattori comuni, cioè la $\boldsymbol{Cov(u_i, F_j) = 0}$, per ogni $\boldsymbol{i}$ e per ogni $\boldsymbol{j}$;
- i **fattori unici** sono ortogonali, non correlano tra loro, cioè la $\boldsymbol{Cov(u_i, u_j) = 0}$, per ogni $\boldsymbol{i}$ e per ogni $\boldsymbol{j}$;
- i **fattori comuni** possono essere correlati tra loro, se indichiamo con la lettera $\boldsymbol{\Phi}$ (*fi*) la matrice di correlazione abbiamo che se i fattori sono indipendenti o ortogonali $\boldsymbol{\Phi = I}$.

ANALISI FATTORIALE

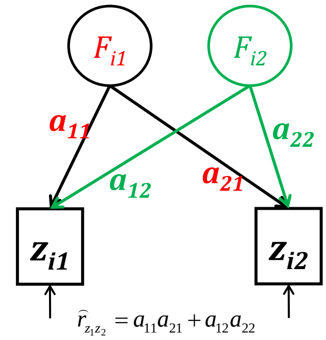
$$z_{ik} = F_{i1}a_{k1} + F_{i2}a_{k2} + \ldots + F_{im}a_{km} + u_{ik}$$

$$R = AA' + U^2$$

Equazione fondamentale dell'AF
(derivabile in base agli assunti)

$$\hat{R} = AA'$$

$$R - \hat{R} = fit$$



ANALISI FATTORIALE

L'analisi **fattoriale** è un metodo statistico sviluppato da Spearman (1904). In generale può essere definita come quella tecnica che consente di esprimere le variazioni di un certo numero di variabili k utilizzando un numero <u>inferiore</u> di **dimensioni** o **costrutti** non osservabili (o latenti) chiamati **fattori** (**$f < k$**). In generale si parte da una matrice di **<u>correlazioni o di covarianza tra le variabili osservate</u>** e si arriva a (**1**) una matrice di **<u>correlazioni tra le variabili osservate e i fattori</u>** (**matrice delle saturazioni**) e (**2**) una matrice di **<u>correlazione o covarianza tra i fattori</u>** (**varianza-covarianza**).

	V1	V2	V3	V4
V1	1	.32	.25	.68
V2	.32	1	.45	.41
V3	.25	.45	1	.64
V4	.68	.41	.64	1

	F1	F2
F1	1	.02
F2	.02	1

	F1	F2
V1	.84	.12
V2	.62	.01
V3	.05	.75
V4	.08	.61

ANALISI FATTORIALE

Passi da seguire per effettuare un'analisi fattoriale:

(1) calcolare la matrice delle correlazioni;

(2) estrarre i fattori;

(3) definire il numero di fattori da mantenere;

(4) calcolare la matrice delle saturazioni;

(5) rotazione della soluzione fattoriale;

(6) interpretazione dei fattori;

(7) calcolo dei punteggi fattoriali.

ANALISI FATTORIALE

L'**analisi fattoriale** si distingue in:

• **Analisi Fattoriale Esplorativa** (**AFE** o ***FA***): quando non si conosce il numero di fattori da estrarre e ciascun indicatore può essere associato a ciascun fattore.

• **Analisi Fattoriale Confermativa** (**AFC** o ***CFA***): quando viene definita *a priori* una struttura di fattori latenti e di relazioni tra fattori e indicatori e la si sottopone a verifica.

Nell'**analisi fattoriale** si distingue tra:

• **Componenti**: i fattori sono derivabili direttamente dai dati (in questo caso la soluzione è unica). **Modello formativo**.

• **Fattori**: i fattori sono teorici e vengono stimati a partire da i dati (per questo le soluzioni sono infinite). **Modello riflettivo**.

ANALISI COMPONENTI PRINCIPALI (ACP o PCA)

Una **componente principale** è una combinazione lineare di variabili espressa dalla seguente equazione:

$$C_i = b_{i1} z_1 + b_{i2} z_2 + b_{i3} z_3 + \ldots + b_{ij} z_j$$

in cui ogni parametro b_{ij} rappresenta il **peso** che la variabile x_i, standardizzata (z_i), ha nel determinare la componente stessa.

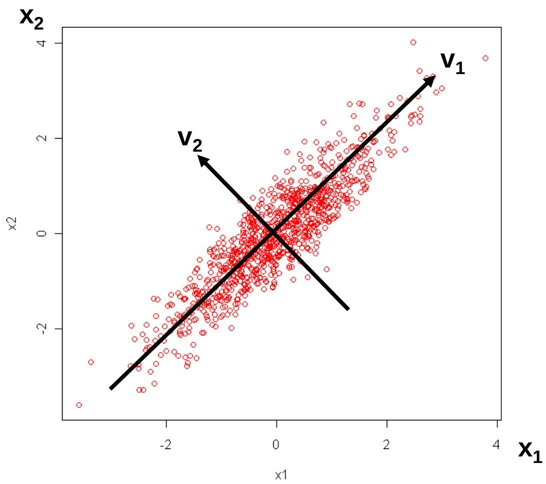
ANALISI COMPONENTI PRINCIPALI

Obiettivo dell'**ACP** è riuscire a ricavare la matrice delle **saturazioni** (A) utile a riprodurre la matrice delle **correlazioni** (R) utilizzando un numero di componenti inferiore al numero delle variabili (k).
Tale risultato viene ottenuto – matematicamente – attraverso l'individuazione di alcuni elementi che caratterizzano le matrici di correlazione, in grado di sintetizzare la varianza e la correlazione tra gli elementi:

1. le radici caratteristiche o ***autovalori*** (L);
2. i vettori associati o ***autovettori*** (V).

$$R = VLV'$$

ANALISI COMPONENTI PRINCIPALI



ANALISI COMPONENTI PRINCIPALI

Gli ***autovalori*** (***L***) di una matrice di correlazione forniscono importanti informazioni sulla varianza delle variabili originali:

• per ogni componente, l'***autovalore*** diviso il numero di variabili è uguale alla **proporzione (%)** di varianza spiegata dalla componente;

• per ogni componente, la somma delle saturazioni elevate al quadrato (colonna) è uguale all'***autovalore***, mentre la somma delle saturazioni al quadrato per variabile (riga) è uguale alla ***comunalità*** della variabile (h^2);

• la somma degli ***autovalori*** è uguale alla somma delle varianze standardizzate delle variabili.

$$R = \begin{bmatrix} 1 & .884 & .617 & .920 \\ .884 & 1 & .720 & .861 \\ .617 & .720 & 1 & .693 \\ .920 & .861 & .693 & 1 \end{bmatrix}$$

Matrice delle correlazioni

$$\sum r_{i1} = 3.421; \sum r_{i2} = 3.465; \ldots$$

Calcolo del vettore di prova (Ua1): somma per colonna.

$$\sqrt{\sum 3.421^2 + 3.465^2 + 3.030^2 + 3.475^2} = 6.706$$

Calcolo 1° autovalore di prova.

$$\frac{3.421}{6.706} = .510; \frac{3.465}{6.706} = .517; \frac{3.030}{6.706} = .452; \frac{3.475}{6.706} = .518$$

Normalizzo il vettore di prova (Ua1): Va1.

$$\begin{bmatrix} 1 & .884 & .617 & .920 \\ .884 & 1 & .720 & .861 \\ .617 & .720 & 1 & .693 \\ .920 & .861 & .693 & 1 \end{bmatrix} \times \begin{bmatrix} .510 \\ .517 \\ .452 \\ .518 \end{bmatrix} = \begin{bmatrix} 1.723 \\ 1.739 \\ 1.498 \\ 1.746 \end{bmatrix}$$

Calcolo del vettore di prova (Ua2): prodotto di R e Va1.

$$(1 \times .510) + (.884 \times .517) + (.617 \times .452) + (.920 \times .518) = 1.723$$

$$\ldots$$

$$\sqrt{\sum 1.723^2 + 1.739^2 + 1.498^2 + 1.746^2} = 3.359$$

Calcolo il 2° autovalore di prova.

$$\frac{1.723}{3.359} = .513; \frac{1.739}{3.359} = .518; \frac{1.498}{3.359} = .446; \frac{1.746}{3.359} = .520$$

Normalizzo il vettore di prova (Ua2): Va2. Da confrontare con Va1.

$$\begin{bmatrix} .510 \\ .517 \\ .452 \\ .518 \end{bmatrix}$$

$$\begin{bmatrix} 1 & .884 & .617 & .920 \\ .884 & 1 & .720 & .861 \\ .617 & .720 & 1 & .693 \\ .920 & .861 & .693 & 1 \end{bmatrix} \times \begin{bmatrix} .513 \\ .518 \\ .446 \\ .520 \end{bmatrix} = \begin{bmatrix} 1.725 \\ 1.740 \\ 1.496 \\ 1.747 \end{bmatrix}$$

Calcolo del vettore di prova (Ua3): prodotto di R e Va2.

$$(1 \times .513) + (.884 \times .518) + (.617 \times .446) + (.920 \times .520) = 1.725$$

$$\ldots$$

$$\sqrt{\sum 1.724^2 + 1.740^2 + 1.496^2 + 1.747^2} = 3.360$$ ← Calcolo il 3° autovalore di prova.

$$\frac{1.724}{3.360} = .513; \frac{1.740}{3.360} = .518; \frac{1.496}{3.360} = .445; \frac{1.747}{3.360} = .520$$ ← Normalizzo il vettore di prova (Ua3): Va3. Da confrontare con Va2.

$$\begin{array}{ccccccc} Va1 & & Va2 & & Va3 & & Va4 \\ \begin{bmatrix} .510 \\ .517 \\ .452 \\ .518 \end{bmatrix} & \neq & \begin{bmatrix} .513 \\ .518 \\ .446 \\ .520 \end{bmatrix} & \neq & \begin{bmatrix} .513 \\ .518 \\ .445 \\ .520 \end{bmatrix} & = & \begin{bmatrix} .513 \\ .518 \\ .445 \\ .520 \end{bmatrix} \end{array}$$

→ **stop**

$$A = \begin{bmatrix} .513 \\ .518 \\ .445 \\ .520 \end{bmatrix} \times \sqrt{3.360} = \begin{bmatrix} .940 \\ .949 \\ .816 \\ .953 \end{bmatrix}$$ ← Calcolo le saturazioni fattoriali.

$$\hat{R} = \begin{bmatrix} .884 .839 .767 .896 \\ .893 .902 .775 .905 \\ .767 .775 .665 .778 \\ .896 .905 .778 .909 \end{bmatrix}$$ ← Calcolo la matrice delle correlazioni residua. Moltiplicando le saturazioni relative a ciascuna coppia di variabili.

$$\sum \hat{r}_{i1} =$$ ← Inizio la procedura per l'estrazione della 2° componente ortogonale.

ANALISI FATTORIALE

L'analisi delle componenti principali (**ACP**) consente di **stimare** la matrice delle saturazioni. Tuttavia, nell'estrazione delle **CP** viene presa in considerazione sia la varianza comune (covarianza) sia la varianza unica. Per questo sono state proposte diverse tecniche di analisi fattoriale (**AF**), o di "**estrazione**", che cercano di prendere in considerazione esclusivamente la covarianza (**comunalità**) tra le variabili separandola dalla parte unica (**unicità**). La differenza tra le tecniche di **AF** riguarda la procedura adottata e l'obiettivo dell'analisi. Le principali sono:

(•) **analisi dei fattori principali** o *principal axis factoring* (**PAF**);

(•) **minimi quadrati** o *minimum residual* (**MR**)

(•) **massima verosimiglianza** o *maximum likelihood* (**ML**)

PRINCIPAL AXIS FACTORING (PAF)

La **PAF** è un metodo di estrazione dei fattori che da un punto di vista matematico risulta estremamente simile all'analisi delle CP. La principale differenza è che nella **PAF** non si utilizza la matrice delle correlazioni completa, ma quella ridotta ($\mathbf{R}_h$). Vale a dire una matrice che nella diagonale principale presenta delle **stime della comunalità** (indeterminatezza). In questo modo si analizza esclusivamente la varianza dei fattori comuni.

	x1	x2	x3	x4	x5	x6
x1	1.00	0.68	0.66	0.24	0.26	0.11
x2	0.68	1.00	0.43	0.13	0.23	0.03
x3	0.66	0.43	1.00	0.24	0.24	0.11
x4	0.24	0.13	0.24	1.00	0.69	0.68
x5	0.26	0.23	0.24	0.69	1.00	0.42
x6	0.11	0.03	0.11	0.68	0.42	1.00

- correlazione max
- triade $(r_{AB} \times r_{AC})/r_{BC}$
- la correlazione media
- il coefficiente $\mathbf{R}^2$

In ogni caso la stima è solo iniziale. Il valore viene sostituito fino all'individuazione di un valore stabile.

MINIMUM RESIDUAL (MR)

In questo metodo i fattori vengono estratti in modo da minimizzare **la differenza al quadrato** tra gli elementi della matrice delle correlazioni osservate ($\mathbf{R_1}$) (gli elementi fuori dalla diagonale) e quelle riprodotte utilizzando i fattori estratti ($\mathbf{R'_1}$).

$$\min = \sum (R_1 - \hat{R}_1)^2 \qquad \hat{R}_1 = AA'$$

Minimi quadrati ordinari (*Unweighted Least Squares*, ULS): tutte le variabili hanno lo stesso peso.

Minimi quadrati generalizzati (*Generalized Least Squares*, GLS): le variabili con una maggiore unicità pesano di meno.

La soluzione è simile alla **PAF**.

MAXIMUM LIKELIHOOD (ML)

Mediante la funzione di **massima verosimiglianza** (**ML**) si stimano quelle saturazioni che rendono massima la probabilità di osservare dalla popolazione la matrice campionaria $\mathbf{R_1}$, vale a dire che minimizzino la quantità $\mathbf{R_1} - \mathbf{R'_1}$.

Come nella procedura MR si considerano esclusivamente gli elementi fuori dalla diagonale. Quindi, non sono necessarie stime della **comunalità**, ma è necessario indicare il **numero di fattori**. La procedura segue un **processo iterativo** fino a quando la funzione di verosimiglianza non subisce minimizzazioni rilevanti.

L'assunto di base è che le variabili abbiano una **distribuzione normale multivariata** e che un modello che prevede ***k*-fattori** sia vero nella popolazione.

Se si applica il metodo **ML** o il metodo **GLS** è possibile avere un indice di *fit* (χ^2) per valutare l'**adeguatezza della soluzione fattoriale**.

NUMERO DI FATTORI

Il numero di fattori o componenti da estrarre è la decisione più importante da prendere nell'AFE. Diversi criteri sono stati proposti:

- ***minimo autovalore*** (***mineigen***), estrarre tutti i fattori che hanno un autovalore maggiore o uguale a 1;
- **grafico degli autovalori** (***scree test***), guardando il grafico degli autovalori si sceglie il numero di fattori corrispondente al fattore dopo il quale la curva diventa piatta;
- **analisi parallela** (***parallel analysis***), utilizzando procedure di simulazione si generano dati casuali e si estraggono i fattori, la regola prevede di scegliere solo quei fattori che hanno degli autovalori superiori a quelli casuali;
- **bontà dell'adattamento** (***goodness of fit***), ove possibile (es. ML & GLS) valutazione degli indici di *fit*.
- **percentuale di varianza spiegata**, valutazione della percentuale di varianza spiegata dal fattore, che dovrebbe non essere inferiore a un valore **minimo** (es. 10%, 5% o 1%), o dalla soluzione (es. 80%);
- **massima correlazione residua**, analizzare la matrice delle correlazioni residue, considerando adeguate soluzioni con valori inferiori a |.10|;
- **cross-validazione** (***cross-validation***), dividere il campione in due metà stimando il modello su un campione e verificandolo sull'altro (indice di *cross-validation* [CVI = 0]);
- **interpretabilità della soluzione**, verificare l'interpretabilità della soluzione mediante l'analisi delle saturazioni (struttura semplice).

INTERPRETAZIONE DEI FATTORI

In generale è possibile affermare che le variabili che presentano saturazioni superiori a **|.30|** (**9%**) possono essere considerate dei ***marker*** del fattore. Vale a dire possono essere utilizzate per interpretare il significato del fattore. La variabile con la **saturazione più elevata** è quella più rilevante per il fattore considerato.

Tuttavia, per un'interpretazione univoca, è necessario che la stessa variabile non abbia saturazioni superiori a **|.30|** su altri fattori o un rapporto tra le due più alte saturazioni superiore a **2** (**struttura semplice**). In caso contrario si dice che la variabile è bifattoriale.

Dal momento che l'esito di un'AFE non necessariamente porta a una soluzione facilmente interpretabile è utile applicare delle trasformazioni matematiche che consentano di giungere a una **struttura fattoriale semplice** che renda interpretabili in modo univoco i fattori. Tali procedure sono le **rotazioni fattoriali**.

ANALISI FATTORIALE

Thurstone (1947) ha introdotto il concetto di **struttura semplice** come obiettivo della **rotazione** per poter interpretare il significato dei fattori. Per questo ha proposto cinque criteri:

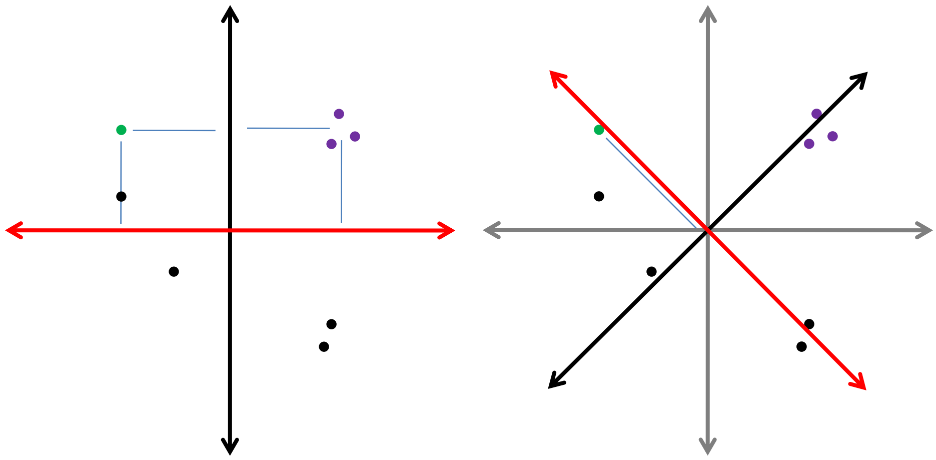
1. ogni riga della matrice ruotata dovrebbe contenere uno **0**;
2. in ogni fattore il numero minimo delle saturazioni **0** dovrebbe essere pari al numero di fattori;
3. per ogni coppia di fattori ci dovrebbero essere variabili con saturazioni **0** su un fattore e significative sull'altro (**>.30**);
4. per ogni coppia di fattori un'ampia proporzione delle saturazioni dovrebbe essere **0**;
5. per ogni coppia di fattori ci dovrebbero essere solo poche variabili con saturazioni significative su entrambi i fattori (**>.30**).

ANALISI FATTORIALE

ITEM	F1	F2
X1	.6	.5
X2	.9	.4
X3	.8	.7
Y1	.3	.4
Y2	.6	.9
Y3	.5	.8

ITEM	F1	F2
X1	1	0
X2	1	0
X3	1	0
Y1	0	1
Y2	0	1
Y3	0	1

ROTAZIONE ORTOGONALE



ANALISI FATTORIALE

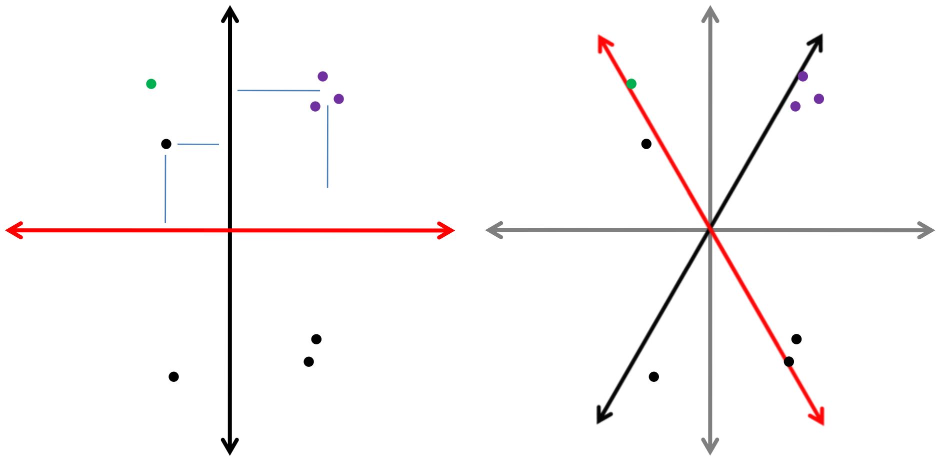
In alcuni casi la rotazione prescelta può essere di tipo obliquo, e i fattori risultanti risulteranno quindi correlati.

Ciò può essere opportuno qualora sia teoricamente giustificato attendersi una correlazione fra i costrutti indagati dalle variabili osservate.

Nella **rotazione obliqua** bisogna distinguere fra due tipi di matrice delle saturazioni:

1. **matrice di struttura**: che riporta le correlazioni fra variabili e fattori (non considerando le correlazioni tra le variabili osservate);
2. **matrice dei *pattern*** (configurazione o "matrice dei modelli") che riporta la relazione specifica di ciascuna variabile con il fattore (coefficienti **parziali**).

ROTAZIONE OBLIQUA



ANALISI FATTORIALE

[assunti e prerequisiti]

• Presenza di strutture rilevanti nei dati (es. nella matrice ***R*** almeno alcune correlazioni >|.30|);

• test di sfericità di Bartlett significativo;

• test di adeguatezza campionaria di Kayser-Meyer-Olkin (*KMO* > .60);

• presenza di variabili misurate su scala almeno a intervalli (se Likert con almeno 5 livelli) e distribuite normalmente (Pearson), altrimenti andrebbero adottate tecniche sviluppate per scale non parametriche;

• assenza di osservazioni ***outlier*** o di variabili non correlate con le altre in analisi;

• adeguata ampiezza del campione (*ss* > 300);

• adeguato rapporto (***odds***) tra variabili e fattori latenti (> 3).

ANALISI FATTORIALE
[regole pratiche]

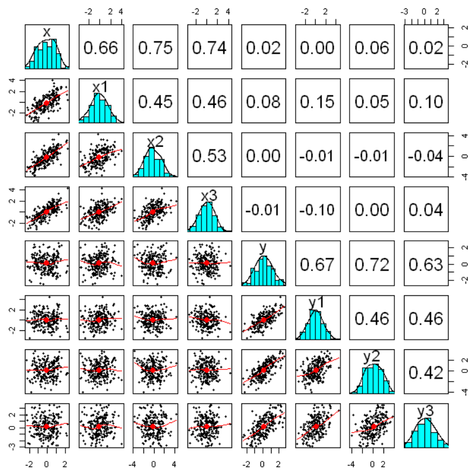
- almeno 4 variabili per definire il significato di un fattore;
- almeno 100 ss, ma se le comunalità sono < .40 anche 400;
- adeguata variabilità nel campione;
- attenzione alla differenza tra ACP e AF, preferendo l'estrazione ML se possibile;
- usare diversi metodi per la scelta del n. di fattori;
- preferire la rotazione obliqua (***oblimin*** o ***promax***).
- usare l'AFC per la verifica di modelli teorici;

ACP
[Db & descrittive]

```
> library(psych)
> library(GPArotation)
> set.seed(10)
> x<-rnorm(200); y<-rnorm(200);
> x1<-x+rnorm(200); x2<-x+rnorm(200); x3<-x+rnorm(200);
> y1<-y+rnorm(200); y2<-y+rnorm(200); y3<-y+rnorm(200);
> prova<-as.data.frame(cbind(x,x1,x2,x3,y,y1,y2,y3))

> describe(prova)
   var   n  mean   sd median trimmed  mad   min  max range  skew kurtosis   se
x    1 200 -0.12 0.95  -0.14   -0.10 1.14 -2.32 2.22  4.54 -0.08    -0.73 0.07
x1   2 200 -0.11 1.48  -0.14   -0.10 1.53 -3.50 4.06  7.57 -0.01    -0.20 0.10
x2   3 200 -0.20 1.47  -0.26   -0.21 1.63 -3.69 4.23  7.92  0.11    -0.24 0.10
x3   4 200  0.01 1.29   0.05    0.02 1.34 -3.23 4.31  7.54  0.01    -0.01 0.09
y    5 200  0.11 1.01   0.08    0.10 1.01 -2.44 2.64  5.08  0.09    -0.22 0.07
y1   6 200  0.06 1.41  -0.03    0.04 1.31 -3.55 4.54  8.09  0.11    -0.05 0.10
y2   7 200  0.13 1.51   0.10    0.14 1.74 -3.99 3.63  7.62 -0.05    -0.56 0.11
y3   8 200  0.19 1.43   0.11    0.18 1.51 -2.89 3.48  6.37  0.12    -0.48 0.10

> paf.2<-paf(as.matrix(prova), eigcrit=1, convcrit=.001)
> paf.2$KMO
[1] 0.701
> bartlett.test(prova)

        Bartlett test of homogeneity of variances

data:  prova
Bartlett's K-squared = 80.757, df = 7, p-value = 9.656e-15
```

ACP
[correlazioni]

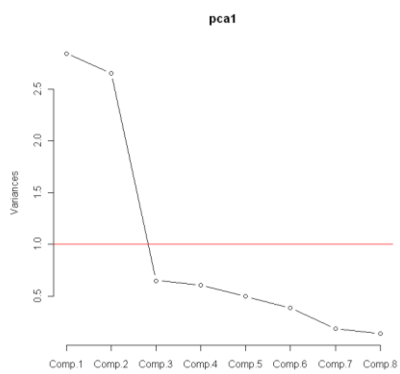
```
> corr.test(prova)
Call:corr.test(x = prova)
Correlation matrix
      x   x1    x2    x3     y    y1    y2    y3
x  1.00 0.66  0.75  0.74  0.02  0.00  0.06  0.02
x1 0.66 1.00  0.45  0.46  0.08  0.15  0.05  0.10
x2 0.75 0.45  1.00  0.53  0.00 -0.01 -0.01 -0.04
x3 0.74 0.46  0.53  1.00 -0.01 -0.10  0.00  0.04
y  0.02 0.08  0.00 -0.01  1.00  0.67  0.72  0.63
y1 0.00 0.15 -0.01 -0.10  0.67  1.00  0.46  0.46
y2 0.06 0.05 -0.01  0.00  0.72  0.46  1.00  0.42
y3 0.02 0.10 -0.04  0.04  0.63  0.46  0.42  1.00
Sample Size
     x  x1  x2  x3   y  y1  y2  y3
x  200 200 200 200 200 200 200 200
x1 200 200 200 200 200 200 200 200
x2 200 200 200 200 200 200 200 200
x3 200 200 200 200 200 200 200 200
y  200 200 200 200 200 200 200 200
y1 200 200 200 200 200 200 200 200
y2 200 200 200 200 200 200 200 200
y3 200 200 200 200 200 200 200 200
Probability value
      x   x1   x2   x3    y   y1   y2   y3
x  0.00 0.00 0.00 0.00 0.77 0.96 0.40 0.82
x1 0.00 0.00 0.00 0.00 0.23 0.03 0.46 0.16
x2 0.00 0.00 0.00 0.00 0.98 0.91 0.91 0.59
x3 0.00 0.00 0.00 0.00 0.87 0.14 0.95 0.53
y  0.77 0.23 0.98 0.87 0.00 0.00 0.00 0.00
y1 0.96 0.03 0.91 0.14 0.00 0.00 0.00 0.00
y2 0.40 0.46 0.91 0.95 0.00 0.00 0.00 0.00
y3 0.82 0.16 0.59 0.53 0.00 0.00 0.00 0.00
```



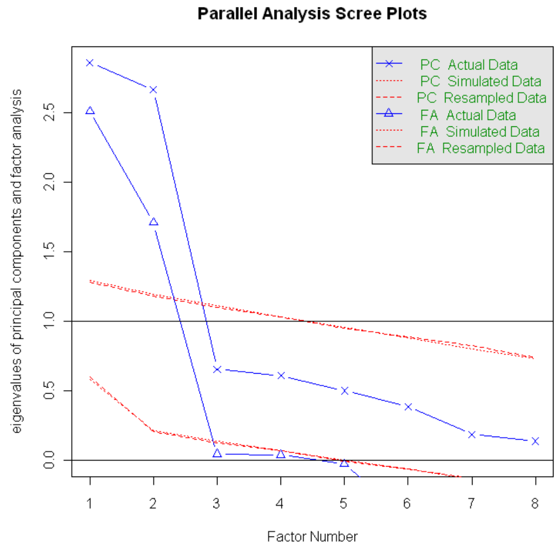
ACP
[valutazione n. CP]

```
> pca1<-princomp(scale(prova))
> summary(pca1)
Importance of components:
                         Comp.1    Comp.2     Comp.3     Comp.4     Comp.5     Comp.6     Comp.7     Comp.8
Standard deviation     1.686160 1.6282097 0.80655658 0.77888275 0.70581224 0.62111800 0.43233641 0.37112148
Proportion of Variance 0.357178 0.3330486 0.08172532 0.07621336 0.06258429 0.04846577 0.02348176 0.01730291
Cumulative Proportion  0.357178 0.6902266 0.77195192 0.84816528 0.91074956 0.95921534 0.98269709 1.00000000
```

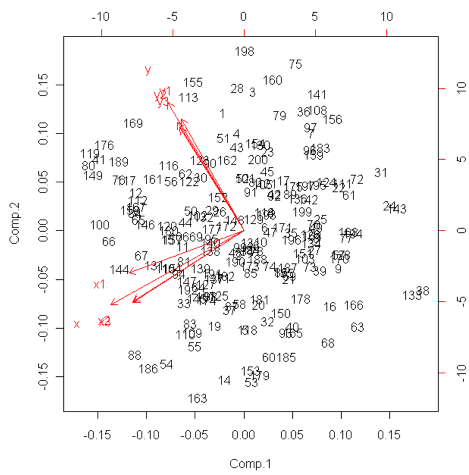
```
> pca1$sdev^2
   Comp.1    Comp.2    Comp.3    Comp.4    Comp.5    Comp.6    Comp.7    Comp.8
2.8431369 2.6510669 0.6505335 0.6066583 0.4981709 0.3857876 0.1869148 0.1377312
```



ACP
[analisi parallela]



ACP
[grafico fattori e ss]

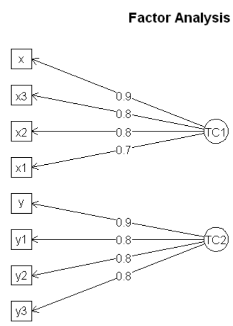


ACP
[saturazioni]

```
> loadings(pca1)

Loadings:
   Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
x  -0.490 -0.283                              0.237  0.786
x1 -0.424 -0.160  0.595        -0.540  0.273 -0.138 -0.234
x2 -0.409 -0.270 -0.110 -0.288  0.638  0.353        -0.360
x3 -0.408 -0.274 -0.345  0.339 -0.136 -0.623        -0.333
y  -0.280  0.490 -0.129                      -0.780  0.222
y1 -0.229  0.426  0.492 -0.268  0.286 -0.524  0.304
y2 -0.245  0.414 -0.500 -0.364 -0.399  0.192  0.395 -0.183
y3 -0.235  0.393         0.768  0.205  0.317  0.238

               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
SS loadings     1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000
Proportion Var  0.125  0.125  0.125  0.125  0.125  0.125  0.125  0.125
Cumulative Var  0.125  0.250  0.375  0.500  0.625  0.750  0.875  1.000
```


ACP
[saturazioni ruotate “*oblimin*”]

```
> pca2 <- principal(prova, nfactors=2, rotate="oblimin", scores=TRUE )
> pca2 # print results
Principal Components Analysis
Call: principal(r = prova, nfactors = 2, rotate = "oblimin", scores = TRUE)
Standardized loadings based upon correlation matrix
     TC1   TC2   h2   u2
x   0.95  0.00 0.90 0.10
x1  0.75  0.12 0.58 0.42
x2  0.82 -0.05 0.67 0.33
x3  0.82 -0.05 0.68 0.32
y   0.00  0.93 0.86 0.14
y1 -0.02  0.80 0.63 0.37
y2  0.01  0.79 0.63 0.37
y3  0.01  0.75 0.57 0.43

               TC1  TC2
SS loadings    2.81 2.71
Proportion Var 0.35 0.34
Cumulative Var 0.35 0.69

 With factor correlations of
     TC1  TC2
TC1 1.00 0.02
TC2 0.02 1.00

Test of the hypothesis that 2 factors are sufficient.

The degrees of freedom for the null model are  28  and the objective function was  4.18
The degrees of freedom for the model are 13  and the objective function was  0.49
The number of observations was  200  with Chi Square =  94.83  with prob <  1.7e-14

Fit based upon off diagonal values = 0.96>
```



ACP

[matrice osservata e residui]

```
> round(cor(prova),3)
        x    x1     x2     x3      y     y1     y2     y3
x   1.000 0.660  0.747  0.741  0.021 -0.003  0.060  0.016
x1  0.660 1.000  0.448  0.458  0.085  0.151  0.052  0.099
x2  0.747 0.448  1.000  0.528  0.002 -0.008 -0.008 -0.039
x3  0.741 0.458  0.528  1.000 -0.011 -0.105  0.005  0.044
y   0.021 0.085  0.002 -0.011  1.000  0.673  0.725  0.626
y1 -0.003 0.151 -0.008 -0.105  0.673  1.000  0.456  0.465
y2  0.060 0.052 -0.008  0.005  0.725  0.456  1.000  0.416
y3  0.016 0.099 -0.039  0.044  0.626  0.465  0.416  1.000
> round(pca2$residual, 3)
        x     x1     x2     x3      y     y1     y2     y3
x   0.100 -0.055 -0.028 -0.037 -0.003 -0.004  0.028 -0.017
x1 -0.055  0.418 -0.162 -0.153 -0.046  0.055 -0.068 -0.018
x2 -0.028 -0.162  0.329 -0.146  0.026  0.030  0.003 -0.031
x3 -0.037 -0.153 -0.146  0.324  0.019 -0.061  0.021  0.058
y  -0.003 -0.046  0.026  0.019  0.136 -0.066 -0.011 -0.075
y1 -0.004  0.055  0.030 -0.061 -0.066  0.367 -0.174 -0.135
y2  0.028 -0.068  0.003  0.021 -0.011 -0.174  0.373 -0.182
y3 -0.017 -0.018 -0.031  0.058 -0.075 -0.135 -0.182  0.431
```

Attendibilità

[alpha: α]

```
> alpha(prova[,1:4])

Reliability analysis
Call: alpha(x = prova[, 1:4])

  raw_alpha std.alpha G6(smc) average_r  mean  sd
      0.83      0.86    0.85       0.6 -0.10 1.1

 Reliability if an item is dropped:
   raw_alpha std.alpha G6(smc) average_r
x       0.73      0.73    0.65      0.48
x1      0.83      0.86    0.83      0.67
x2      0.80      0.83    0.80      0.62
x3      0.79      0.83    0.81      0.62

 Item statistics
     n    r r.cor r.drop   mean   sd
x  200 0.94  0.95   0.89 -0.116 0.95
x1 200 0.77  0.65   0.58 -0.106 1.48
x2 200 0.81  0.74   0.64 -0.196 1.47
x3 200 0.82  0.74   0.65  0.013 1.29
```

```
> alpha(prova[,5:8])
Reliability analysis
Call: alpha(x = prova[, 5:8])

  raw_alpha std.alpha G6(smc) average_r mean  sd
      0.82      0.84    0.82      0.56 0.12 1.1

 Reliability if an item is dropped:
   raw_alpha std.alpha G6(smc) average_r
y       0.71      0.71    0.62      0.45
y1      0.78      0.81    0.78      0.59
y2      0.79      0.81    0.76      0.59
y3      0.80      0.83    0.80      0.62

 Item statistics
     n    r r.cor r.drop  mean  sd
y  200 0.92  0.93   0.85 0.114 1.0
y1 200 0.79  0.69   0.61 0.057 1.4
y2 200 0.79  0.72   0.60 0.129 1.5
y3 200 0.77  0.64   0.57 0.194 1.4
```

Regressione (lm)
[punteggi fattoriali]

```
> prova$F1 <- pca2$scores[,1]
> prova$F2 <- pca2$scores[,2]
> attach(prova)
> summary(lm(y~1+F1))

Call:
lm(formula = y ~ 1 + F1)

Residuals:
     Min       1Q   Median       3Q      Max
-2.55070 -0.67222 -0.03066  0.68007  2.53241

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.113795   0.071513   1.591    0.113
F1          -0.001286   0.071670  -0.018    0.986

Residual standard error: 1.011 on 198 degrees of freedom
Multiple R-squared: 1.625e-06,  Adjusted R-squared: -0.005049
F-statistic: 0.0003218 on 1 and 198 DF,  p-value: 0.9857
```

AF
[estrazione "*paf*" & "*oblimin*"]

```
> fa(prova, nfactors=2, rotate="oblimin", fm="pa")
Factor Analysis using method =  pa
Call: fa(r = prova, nfactors = 2, rotate = "oblimin", fm = "pa")
Standardized loadings based upon correlation matrix
     PA1   PA2   h2      u2
x   1.03  0.00 1.06 -0.0632
x1  0.63  0.11 0.41  0.5859
x2  0.72 -0.03 0.52  0.4776
x3  0.73 -0.04 0.53  0.4732
y   0.00  1.00 1.00 -0.0037
y1 -0.02  0.69 0.47  0.5267
y2  0.02  0.69 0.48  0.5207
y3  0.01  0.63 0.40  0.6009

                PA1  PA2
SS loadings    2.51 2.37
Proportion Var 0.31 0.30
Cumulative Var 0.31 0.61

 With factor correlations of
     PA1  PA2
PA1 1.00 0.03
PA2 0.03 1.00

Tucker Lewis Index of factoring reliability =  0.966
RMSEA index =  0.071  and the 90 % confidence intervals are  0.07 0.076
BIC =  -43.47
Fit based upon off diagonal values = 0.99
```

```
WARNING, the factor score fit indices suggest that the solution is degenerate. Try a different method of factor extraction.
```

AF
[estrazione “*minres*” & “*oblimin*”]

```
> > fa(prova, nfactors=2, rotate="oblimin", fm="minres")
Factor Analysis using method =  minres
Call: fa(r = prova, nfactors = 2, rotate = "oblimin", fm = "minres")
Standardized loadings based upon correlation matrix
     MR1   MR2   h2    u2
x   1.00  0.00 1.00 0.005
x1  0.66  0.07 0.44 0.558
x2  0.75 -0.02 0.56 0.439
x3  0.74 -0.03 0.55 0.448
y  -0.01  0.99 0.98 0.015
y1 -0.02  0.68 0.46 0.540
y2  0.04  0.73 0.53 0.467
y3  0.00  0.63 0.40 0.603

                MR1  MR2
SS loadings    2.55 2.38
Proportion Var 0.32 0.30
Cumulative Var 0.32 0.62

 With factor correlations of
     MR1  MR2
MR1 1.00 0.03
MR2 0.03 1.00
```

```
Tucker Lewis Index of factoring reliability =  0.963
RMSEA index =  0.074  and the 90 % confidence intervals are  0.073 0.079
BIC =  -42.4
Fit based upon off diagonal values = 0.99
```

AF
[estrazione “*wls*” & “*oblimin*”]

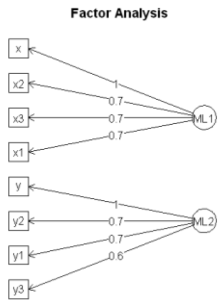
```
> fa(prova, nfactors=2, rotate="oblimin", fm="wls")
Factor Analysis using method =  wls
Call: fa(r = prova, nfactors = 2, rotate = "oblimin", fm = "wls")
Standardized loadings based upon correlation matrix
    WLS1  WLS2   h2     u2
x   1.01  0.00 1.01 -0.012
x1  0.64  0.11 0.43  0.573
x2  0.73 -0.04 0.54  0.463
x3  0.74 -0.04 0.54  0.459
y   0.00  0.99 0.99  0.014
y1 -0.02  0.69 0.47  0.526
y2  0.01  0.70 0.49  0.507
y3  0.01  0.63 0.40  0.597

               WLS1 WLS2
SS loadings    2.50 2.37
Proportion Var 0.31 0.30
Cumulative Var 0.31 0.61

 With factor correlations of
     WLS1 WLS2
WLS1 1.00 0.03
WLS2 0.03 1.00
```

```
Tucker Lewis Index of factoring reliability =  0.961
RMSEA index =  0.076  and the 90 % confidence intervals are  0.075 0.081
BIC =  -41.59
Fit based upon off diagonal values = 0.99
```

```
WARNING, the factor score fit indices suggest that the solution is degenerate. Try a different method of factor extraction.
```

AF
[estrazione "*ml*" & "*oblimin*"]

```
> fa(prova, nfactors=2, rotate="oblimin", fm="ml", scores=TRUE)
Factor Analysis using method =  ml
Call: fa(r = prova, nfactors = 2, rotate = "oblimin", scores = TRUE,
    fm = "ml")
Standardized loadings based upon correlation matrix
     ML1   ML2   h2    u2
x    1.00  0.00 1.00 0.005
x1   0.66  0.07 0.44 0.559
x2   0.75 -0.02 0.56 0.439
x3   0.74 -0.03 0.55 0.448
y   -0.01  1.00 1.00 0.005
y1  -0.02  0.67 0.46 0.545
y2   0.04  0.73 0.53 0.471
y3   0.00  0.63 0.39 0.607

                ML1  ML2
SS loadings    2.55 2.38
Proportion Var 0.32 0.30
Cumulative Var 0.32 0.62

 With factor correlations of
     ML1  ML2
ML1 1.00 0.03
ML2 0.03 1.00
```




```
Tucker Lewis Index of factoring reliability =  0.964
RMSEA index =  0.073  and the 90 % confidence intervals are  0.073 0.079
BIC =  -42.64
Fit based upon off diagonal values = 0.99
```

AF
residui ["*ml*" & "*oblimin*"]

```
> round(ml.2$residual, 3)
       x     x1     x2     x3      y     y1     y2     y3
x  0.005  0.001  0.001  0.000  0.000  0.000  0.001  0.000
x1 0.001  0.559 -0.046 -0.030 -0.001  0.105 -0.039  0.043
x2 0.001 -0.046  0.439 -0.029  0.001  0.004 -0.042 -0.042
x3 0.000 -0.030 -0.029  0.448  0.000 -0.084 -0.019  0.049
y  0.000 -0.001  0.001  0.000  0.005  0.000  0.000  0.000
y1 0.000  0.105  0.004 -0.084  0.000  0.545 -0.033  0.042
y2 0.001 -0.039 -0.042 -0.019  0.000 -0.033  0.471 -0.040
y3 0.000  0.043 -0.042  0.049  0.000  0.042 -0.040  0.607
```