

COMIC – Guida all'uso

(Ver. 1.0 – anno 2017)

Pierpaolo Massoli

Adriano Pareto

Istat

Indice¹

1. Finalità del software
2. Requisiti del software
3. Cosa fa il software?
4. Metodi implementati
5. Struttura delle directory
6. I dati di input
 - L'interfaccia utente*
7. Gli output
 - L'indice dei risultati*
 - Grafici e tabelle*
8. Sviluppi futuri

Riferimenti bibliografici

¹ I paragrafi 2, 3, 5 e 8 sono redatti da P. Massoli; i paragrafi 1, 4, 6 e 7 sono redatti da A. Pareto.

1. Finalità del software

COMIC (COMposite Indices Creator) è un software sviluppato presso l'Istituto Nazionale di Statistica (Istat), nell'ambito del gruppo BES (Benessere Equo e Sostenibile).

Il software consente di:

- calcolare indici sintetici e costruire graduatorie;
- comparare agevolmente diversi metodi di sintesi per scegliere il più opportuno;
- scrivere una reportistica efficace dei risultati ottenuti.

Com'è noto, un indice sintetico (o indice composito) è una combinazione matematica (o aggregazione) di un insieme di indicatori elementari² (variabili) che rappresentano le diverse componenti di un concetto multidimensionale da misurare (per es., sviluppo, qualità della vita, benessere, ecc.). Quindi, gli indici sintetici sono usati per misurare concetti che non possono essere catturati da un unico indicatore.

Generalmente, un indice sintetico dovrebbe essere basato su un quadro teorico che consenta di selezionare, combinare e pesare gli indicatori elementari in modo da riflettere le dimensioni o la struttura del fenomeno che si sta misurando. Tuttavia, la sua costruzione non è semplice e, spesso, richiede una serie di decisioni/scelte (metodologiche o no) da prendere.

La procedura per costruire un indice sintetico prevede i seguenti passi.

1) *Definizione del fenomeno da misurare.* La definizione del concetto dovrebbe fornire un senso chiaro di ciò che si intende misurare con l'indice sintetico. Essa dovrebbe riferirsi a un quadro teorico, comprendente diversi sottogruppi e indicatori sottostanti. Un aspetto fondamentale riguarda l'identificazione del modello di misurazione, per il quale si distinguono due diversi approcci:

- modello *riflessivo*, se gli indicatori sono visti come 'effetto' del fenomeno da misurare, per cui un cambiamento nella variabile latente si riflette in un cambiamento degli indicatori osservati (gli indicatori sono intercambiabili e le correlazioni tra di essi sono spiegate dal modello);
- modello *formativo*, se gli indicatori sono visti come 'causa' del fenomeno da misurare, per cui un cambiamento nella variabile latente non implica necessariamente un

² Un indicatore elementare è un dato 'elaborato' costruito, generalmente, rapportando un dato 'grezzo' ad un altro che ne costituisce una base di riferimento (per es., "reddito pro-capite").

cambiamento di tutti gli indicatori osservati (gli indicatori non sono intercambiabili e le correlazioni tra di essi non sono spiegate dal modello).

- 2) *Selezione di un gruppo di indicatori elementari.* La forza e la debolezza di un indice sintetico riflettono la qualità degli indicatori elementari sottostanti. Gli indicatori dovrebbero essere selezionati in base alla loro rilevanza, validità, tempestività, disponibilità, ecc. La fase di selezione è il risultato di un compromesso tra possibili ridondanze e perdita di informazione. Un approccio statistico alla scelta degli indicatori consiste nel calcolare le correlazioni tra potenziali indicatori e includere quelli meno correlati tra loro. Tuttavia, il processo di selezione dipende dal modello di misurazione adottato: in un modello *riflessivo*, tutti gli indicatori devono essere correlati tra loro, mentre in un modello *formativo* possono essere incorrelati.
- 3) *Normalizzazione degli indicatori elementari.* La normalizzazione ha lo scopo di rendere gli indicatori comparabili in quanto essi, spesso, sono espressi in unità di misura diverse e possono avere polarità³ differenti. Pertanto, è necessario portare gli indicatori a uno stesso standard, invertendo la polarità, laddove necessario, e trasformandoli in numeri puri, adimensionali. Esistono vari metodi di normalizzazione, come la trasformazione in indici relativi (o metodo Min-Max) e la standardizzazione (calcolo dei z-scores).
- 4) *Aggregazione degli indicatori normalizzati.* E' la combinazione di tutte le componenti per formare l'indice sintetico (funzione matematica). Tale passo richiede la definizione dell'importanza di ciascun indicatore elementare (sistema di ponderazione) e l'identificazione della tecnica di sintesi (compensativa o non-compensativa). Il sistema più semplice e usato per la definizione del sistema di ponderazione – ma non per questo esente da critiche – consiste nell'assegnare lo stesso peso a tutti gli indicatori. Per quanto riguarda la tecnica di sintesi, si distinguono due approcci:
 - approccio *compensativo*, se gli indicatori elementari sono considerati sostituibili⁴; in tal caso, si adottano delle funzioni lineari, come la media aritmetica;
 - approccio *non compensativo*, se gli indicatori elementari sono considerati non-sostituibili⁵; in tal caso, si adottano delle funzioni non lineari in cui si tiene conto – implicitamente o esplicitamente – dello sbilanciamento tra i diversi valori, in termini di penalizzazione.

³ La 'polarità' (o 'verso') di un indicatore elementare è il segno della relazione tra l'indicatore e il fenomeno da misurare (per es., nella costruzione di un indice sintetico di sviluppo, la "speranza di vita" ha polarità positiva, mentre la "mortalità infantile" ha polarità negativa).

⁴ Gli indicatori elementari sono detti 'sostituibili' se un deficit in un indicatore può essere compensato da un surplus in un altro (per es., un valore basso in "Percentuale di persone che hanno partecipato ad attività spirituali o religiose" può essere compensato da un valore alto in "Percentuale di persone che hanno partecipato a incontri di associazioni ricreative o culturali" e viceversa).

⁵ Gli indicatori elementari sono detti 'non-sostituibili' se un deficit in un indicatore non può essere compensato da un surplus in un altro (per es., un valore basso in "Letti di ospedale per 1.000 abitanti" non può essere compensato da un valore alto in "Medici per 1.000 abitanti" e viceversa).

5) *Validazione dell'indice sintetico*. Consiste nel verificare che l'indice sintetico è coerente con il quadro teorico generale. In particolare, occorre valutare la capacità dell'indice di produrre risultati stabili e corretti (Analisi di Influenza e/o Analisi di Robustezza) e la sua capacità discriminante.

2. Requisiti del software

COMIC è scritto interamente in SAS e richiede l'installazione del Base SAS System (ver. 9.1 o superiore). Non sono necessari pacchetti aggiuntivi che potrebbero ostacolare l'impiego.

COMIC è alquanto facile da usare e non richiede alcuna installazione: basta copiare il programma sul proprio pc, indipendentemente dal Sistema Operativo utilizzato (Windows, Linux, Unix, ecc.).

3. Cosa fa il software?

Le principali funzioni di COMIC sono le seguenti:

- elaborazione di dati relativi a qualsiasi tipo di unità statistica, attualmente è ottimizzato per lavorare a livello di regione e ripartizione geografica (macroregione);
- analisi esplorativa dei dati di input (statistiche descrittive, matrice di correlazione, matrix-plot; Analisi in Componenti Principali/Analisi Fattoriale, Scree-plot, ecc.);
- costruzione delle graduatorie delle unità statistiche, per ciascun metodo di sintesi o indice sintetico selezionato;
- creazione di mappe geografiche degli indici sintetici (per regione e ripartizione geografica);
- calcolo delle serie storiche degli indici sintetici (per più anni);
- creazione di grafici “velocità-accelerazione”, per mettere in relazione l'intensità dei valori degli indici sintetici (velocità) con la loro variazione nel tempo (accelerazione);
- confronti tra i vari metodi di sintesi (matrici di cograduazione e scatter-plot)
- Analisi di Influenza (report e grafici);
- Analisi di Robustezza (report e grafici).

4. Metodi implementati

I metodi di sintesi, disponibili su COMIC, sono i seguenti:

- Media indici 0-1;
- Media z-scores;
- Indice di Jevons statico e dinamico;
- Mazziotta-Pareto Index (MPI);
- Adjusted Mazziotta-Pareto Index (AMPI) o MPI corretto;
- Indice media geometrica (IMG).

Media indici 0-1

Data la matrice $\mathbf{X}=\{x_{ijt}\}$ di n righe (unità statistiche), m colonne (indicatori) e p strati (anni), si calcola la matrice trasformata $\mathbf{R}=\{r_{ijt}\}$, con:

$$r_{ijt} = \begin{cases} \frac{(x_{ijt} - \text{Min}_{x_j})}{(\text{Max}_{x_j} - \text{Min}_{x_j})} & \text{se l'indicatore } j \text{ ha polarità positiva;} \\ \frac{(\text{Max}_{x_j} - x_{ijt})}{(\text{Max}_{x_j} - \text{Min}_{x_j})} & \text{se l'indicatore } j \text{ ha polarità negativa;} \end{cases}$$

dove Min_{x_j} e Max_{x_j} sono, rispettivamente, il minimo e il massimo dell'indicatore j , tra le n unità statistiche, nei p anni considerati o dei valori forniti dall'esterno.

L'indice sintetico, per l'unità i al tempo t , è dato dalla formula:

$$M_{it}^{0-1} = \frac{\sum_{j=1}^m r_{ijt}}{m}.$$

Media z-scores

Data la matrice $\mathbf{X}=\{x_{ijt}\}$ di n righe (unità statistiche), m colonne (indicatori) e p strati (anni), si calcola la matrice trasformata $\mathbf{Z}=\{z_{ijt}\}$, con:

$$z_{ijt} = \begin{cases} \frac{(x_{ijt} - M_{x_{jt}})}{S_{x_{jt}}} & \text{se l'indicatore } j \text{ ha polarità positiva;} \\ -\frac{(x_{ijt} - M_{x_{jt}})}{S_{x_{jt}}} & \text{se l'indicatore } j \text{ ha polarità negativa;} \end{cases}$$

dove $M_{x_{jt}}$ e $S_{x_{jt}}$ sono, rispettivamente, la media e lo scostamento quadratico medio dell'indicatore j al tempo t .

L'indice sintetico, per l'unità i al tempo t , è dato dalla formula:

$$M_{it}^z = \frac{\sum_{j=1}^m z_{ijt}}{m}.$$

Indice di Jevons statico e dinamico

Data la matrice $\mathbf{X}=\{x_{ijt}\}$ di n righe (unità statistiche), m colonne (indicatori) e p strati (anni), l'indice sintetico 'statico', per l'unità i al tempo t , è dato dalla formula:

$$JS_{it} = \prod_{j=1}^m \left(\frac{x_{ijt}}{x_{bjt}} 100 \right)^{\frac{1}{m}}$$

dove x_{bjt} è il valore base dell'indicatore j al tempo t (nel caso delle regioni italiane, è la media nazionale nell'anno di riferimento).

L'indice sintetico 'dinamico', per l'unità i al tempo t , è dato dalla formula:

$$JD_{it} = \prod_{j=1}^m \left(\frac{x_{ijt}}{x_{ij(t-1)}} 100 \right)^{\frac{1}{m}}$$

dove $x_{ij(t-1)}$ è il valore dell'indicatore j , per l'unità i al tempo $t-1$.

Mazziotta-Pareto Index (MPI)

Data la matrice $\mathbf{X}=\{x_{ijt}\}$ di n righe (unità statistiche), m colonne (indicatori) e p strati (anni), si calcola la matrice trasformata $\mathbf{Z}=\{z_{ijt}\}$, con:

$$z_{ijt} = \begin{cases} 100 + \frac{(x_{ijt} - M_{x_{jt}})}{S_{x_{jt}}} 10 & \text{se l'indicatore } j \text{ ha polarità positiva;} \\ 100 - \frac{(x_{ijt} - M_{x_{jt}})}{S_{x_{jt}}} 10 & \text{se l'indicatore } j \text{ ha polarità negativa;} \end{cases}$$

dove $M_{x_{jt}}$ e $S_{x_{jt}}$ sono, rispettivamente, la media e lo scostamento quadratico medio dell'indicatore j al tempo t .

L'indice sintetico, per l'unità i al tempo t , è dato dalla formula⁶:

$$\text{MPI}_{it}^{+/-} = M_{z_{it}} \pm S_{z_{it}} \text{cv}_{it}$$

$$\text{dove } M_{z_{it}} = \frac{\sum_{j=1}^m z_{ijt}}{m}; S_{z_{it}} = \sqrt{\frac{\sum_{j=1}^m (z_{ijt} - M_{z_{it}})^2}{m}}; \text{cv}_{it} = \frac{S_{z_{it}}}{M_{z_{it}}}.$$

Adjusted Mazziotta-Pareto Index (AMPI) o MPI corretto

Data la matrice $\mathbf{X}=\{x_{ijt}\}$ di n righe (unità statistiche), m colonne (indicatori) e p strati (anni), si calcola la matrice trasformata $\mathbf{R}=\{r_{ijt}\}$, con:

$$r_{ijt} = \begin{cases} \frac{(x_{ijt} - \text{Min}_{x_{jt}})}{(\text{Max}_{x_{jt}} - \text{Min}_{x_{jt}})} 60 + 70 & \text{se l'indicatore } j \text{ ha polarità positiva;} \\ \frac{(\text{Max}_{x_{jt}} - x_{ijt})}{(\text{Max}_{x_{jt}} - \text{Min}_{x_{jt}})} 60 + 70 & \text{se l'indicatore } j \text{ ha polarità negativa;} \end{cases}$$

dove $\text{Min}_{x_{jt}}$ e $\text{Max}_{x_{jt}}$ sono, rispettivamente, il minimo e il massimo dell'indicatore j , tra le n unità statistiche, nei p anni considerati o dei valori forniti dall'esterno. Tali valori possono essere calcolati in modo da porre uguale a 100 un valore di riferimento (nel caso delle regioni italiane, la media nazionale nell'anno base).

L'indice sintetico, per l'unità i al tempo t , è dato dalla formula¹:

$$\text{MPI}_{it}^{+/-} = M_{r_{it}} \pm S_{r_{it}} \text{cv}_{it}$$

$$\text{dove } M_{r_{it}} = \frac{\sum_{j=1}^m r_{ijt}}{m}; S_{r_{it}} = \sqrt{\frac{\sum_{j=1}^m (r_{ijt} - M_{r_{it}})^2}{m}}; \text{cv}_{it} = \frac{S_{r_{it}}}{M_{r_{it}}}.$$

Indice media geometrica (IMG)

Data la matrice $\mathbf{X}=\{x_{ijt}\}$ di n righe (unità statistiche), m colonne (indicatori) e p strati (anni), si calcola la matrice trasformata $\mathbf{R}=\{r_{ijt}\}$, con:

⁶ E' una forma generalizzata, poiché include due indici in uno. Se l'indicatore è *positivo*, ossia se a variazioni crescenti dell'indicatore corrispondono variazioni positive del fenomeno (per esempio, il benessere), si utilizza MPI⁻. Viceversa, se l'indicatore è *negativo*, ossia se a variazioni crescenti dell'indicatore corrispondono variazioni negative del fenomeno (per esempio, il malessere), si utilizza MPI⁺.

$$r_{ijt} = \begin{cases} \frac{(x_{ijt} - \text{Min}_{x_{jt}})}{(\text{Max}_{x_{jt}} - \text{Min}_{x_{jt}})} 198 + 1 & \text{se l'indicatore } j \text{ ha polarità positiva;} \\ \frac{(\text{Max}_{x_{jt}} - x_{ijt})}{(\text{Max}_{x_{jt}} - \text{Min}_{x_{jt}})} 198 + 1 & \text{se l'indicatore } j \text{ ha polarità negativa;} \end{cases}$$

dove $\text{Min}_{x_{jt}}$ e $\text{Max}_{x_{jt}}$ sono, rispettivamente, il minimo e il massimo dell'indicatore j , tra le n unità statistiche, nei p anni considerati o dei valori forniti dall'esterno. Tali valori possono essere calcolati in modo da porre uguale a 100 un valore di riferimento (nel caso delle regioni italiane, la media nazionale nell'anno base).

L'indice sintetico, per l'unità i al tempo t , è dato dalla formula:

$$\text{IMG}_{it} = \prod_{j=1}^m (r_{ijt})^{\frac{1}{m}}.$$

5. Struttura delle directory

COMIC è distribuito mediante un archivio compresso. Decomprimendo l'archivio, l'albero delle directory è:

- Directory DATI, che contiene le seguenti sub-directory:
 - INPUT: dove possono essere memorizzati i dati di ingresso (in alternativa, è possibile utilizzare una qualsiasi directory selezionata dall'utente);
 - MAPS: dove devono essere memorizzati tutti i dati necessari per creare una mappa geografica;
 - OUTPUT: dove possono essere memorizzati i dati di uscita (in alternativa, è possibile utilizzare una qualsiasi directory selezionata dall'utente);
- Directory PGM, che contiene le macro SAS;
- Directory SETTINGS, in cui sono memorizzati tutti i parametri di esecuzione, allo scopo di facilitare eventuali riesecuzioni del software.

6. I dati di input

COMIC prevede quattro tipi di file di input:

- foglio di lavoro Excel (.xls);
- file di testo ASCII - Comma Separated Values (.csv);
- file di testo ASCII - Tab Separated Values (.txt);
- data set SAS (file senza estensione).

I dati devono essere memorizzati nel file di input, come riportato:

- la prima riga (formato stringa) deve contenere i nomi delle variabili⁷ nella seguente successione: unità statistica (per es., region), anno di riferimento (per es., year), tipologia di unità statistica (per es., type), peso dell'unità statistica (per es., wgt), indicatori elementari (per es., ind1, ind2, ind3, ind4, ecc.);
- dalla seconda riga in poi:
 - la prima colonna (formato stringa) deve contenere le etichette delle unità statistiche;
 - la seconda colonna (formato numerico) deve contenere gli anni di riferimento degli indicatori elementari;
 - la terza colonna (formato numerico) deve contenere la tipologia delle unità statistiche: 1=ripartizione geografica, 2=regione e -1=unità statistica generica;
 - la quarta colonna (formato numerico) deve contenere i pesi (≥ 1) delle unità statistiche; il valore corrispondente alla riga del totale di riferimento viene ignorato.
 - la quinta colonna e le successive (formato numerico) devono contenere i valori degli indicatori elementari;
- alla fine delle righe contenenti le unità statistiche, per ciascun anno, deve essere riportato il totale di riferimento (per es., totale o media nazionale).
- i dati mancanti (missing) non sono ammessi.

Nella figura 1 è riportato un esempio di file di input.

Si noti che alcuni parametri, come la polarità e il peso degli indicatori elementari o i valori minimi e massimi da utilizzare per la loro normalizzazione, non devono essere riportati nel file di input, ma possono essere inseriti nel programma durante l'esecuzione (in caso contrario, si utilizzano le impostazioni di default).

⁷ I nomi delle variabili devono cominciare con una lettera, non devono contenere spazi, caratteri speciali o parole chiave del SAS.

Figura 1

Region	year	type	wgt	ind1	ind2	ind3	ind4
Piemonte	2006	2	1	68,6	7,1	27	10,4
Valle d'Aosta	2006	2	1	70,8	5,4	27	24,2
Liguria	2006	2	1	65,9	8,8	34,1	18,3
Lombardia	2006	2	1	70,4	6,9	25,3	16,7
Bolzano/Bo.	2006	2	1	73,6	3,8	33,5	25,9
Trento	2006	2	1	69,4	6	19	17,6
Veneto	2006	2	1	69,3	7,2	22,5	13,5
Friuli-V.G.	2006	2	1	68,3	7,1	14,9	16,6
Emilia-Rom.	2006	2	1	73,1	6,2	26,4	19,8
Toscana	2006	2	1	68,7	8,7	17,1	16,6
Umbria	2006	2	1	66,9	10,6	25,2	19,3
Marche	2006	2	1	68,2	8,8	30,9	8,5
Lazio	2006	2	1	63,5	14,4	16	23,6
Abruzzo	2006	2	1	62	13,1	17,7	18
Molise	2006	2	1	56,6	20,9	13,4	17,8
Campania	2006	2	1	48,3	29,6	17,2	25,7
Puglia	2006	2	1	49,6	25,8	22,9	22,2
Basilicata	2006	2	1	54,7	23,3	19,6	26,2
Calabria	2006	2	1	49,9	30,1	13,3	36,5
Sicilia	2006	2	1	49,1	29,8	16,2	29,3
Sardegna	2006	2	1	56	21,6	10,2	18,4
Italia	2006	2		62,5	14,5	21,1	20,3
Piemonte	2007	2	1	68,8	7,9	17,5	11,5
Valle d'Aosta	2007	2	1	72,1	5,4	21,8	26,9

L'interfaccia utente

COMIC può essere utilizzato anche da utenti senza alcuna conoscenza di SAS grazie ad una semplice, ma efficace, interfaccia grafica GUI.

Gli utenti comunicano con il software solo attraverso tale interfaccia, implementata attraverso l'istruzione macro SAS Base `%WINDOW/%DISPLAY` per consentire l'inserimento di tutti i parametri necessari all'esecuzione.

I dati immessi tramite la GUI sono salvati in un file ASCII nella directory SETTINGS.

Nella figura 2 è illustrata la prima pagina del menù di input.

In questa pagina occorre selezionare (tramite una x) gli indici che si vogliono calcolare, distinguendo tra 'statici' e 'dinamici'⁸. In quest'ultimo caso, devono essere presenti almeno due anni diversi nel file di input (per es., 2015 e 2016).

Quindi, è necessario specificare se si vuole effettuare l'Analisi di Influenza e/o l'Analisi di Robustezza (tramite y o n).

Infine, occorre inserire il nome del file di input e la directory dei file di output (con i relativi percorsi).

⁸ Gli indici sintetici 'statici' misurano l'intensità del fenomeno in un dato periodo di riferimento (per es., anno), mentre gli indici sintetici 'dinamici' misurano la variazione dell'intensità del fenomeno tra due periodi di riferimento (per es., un anno e quello successivo).

Figura 2

```
Comando ==>
COMIC - COMposite Indices Creator

-----
Lista dei metodi di sintesi disponibili (x per selezionare):

Indici composti STATICI:
Media indici 0-1      X
Media z-scores       X
Jevons statico       X
MPI                  X
MPI corretto          X
IMG                  X

Indici composti DINAMICI:
Jevons dinamico      |
Differenza media indici 0-1
Differenza MPI corretto
Differenza IMG

Effettuare l'analisi di influenza degli indicatori selezionati? (y/n)  X
Effettuare l'analisi di robustezza degli indicatori selezionati? (y/n)

-----
File con gli indicatori base (.xls, .txt, .csv oppure no estensione = data set SAS):
G:\COMIC_SAS9.3\dati\input\esempio1.xls

Directory dei risultati:
G:\COMIC_SAS9.3\dati\output

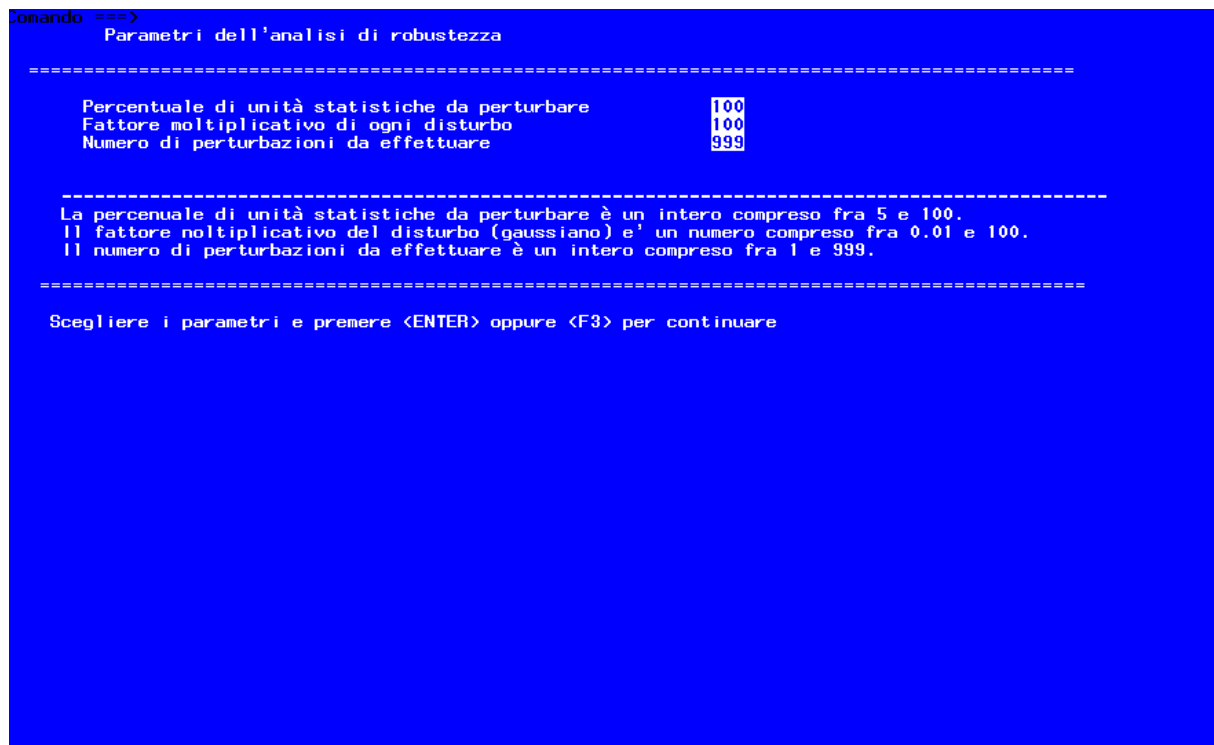
-----
Selezionare uno o più metodi e immettere i percorsi dei file o premere <F3> per continuare
```

L'Analisi di Influenza ha lo scopo di valutare se - e con quale intensità - la graduatoria delle unità statistiche cambia, a seguito dell'eliminazione di un indicatore elementare dall'insieme di partenza. Dati m indicatori elementari, si conducono m replicazioni eliminando, ogni volta, un indicatore diverso e calcolando i valori degli indici sintetici sulla base degli $m-1$ indicatori rimanenti. Per ogni replicazione, si costruiscono le graduatorie delle unità statistiche secondo i vari metodi e, per ciascuna unità, si calcolano le differenze assolute di rango tra la posizione nella graduatoria originaria e la posizione nella graduatoria relativa agli $m-1$ indicatori elementari. Infine, per ciascun indice sintetico, si calcolano la media e lo scostamento quadratico medio delle differenze assolute di rango (shift). L'analisi viene ripetuta per ciascun anno, presente nel file di input.

L'Analisi di Robustezza si propone di verificare la 'stabilità' di un indice sintetico al variare di perturbazioni casuali della matrice di partenza. Se si sceglie di effettuare questa analisi, comparirà una pagina aggiuntiva (figura 3) in cui si chiede l'inserimento dei seguenti parametri:

1. percentuale di unità statistiche da perturbare p (valore compreso tra 5 e 100);
2. fattore moltiplicativo o intensità dei disturbi λ (valore compreso tra 0,01 e 100);
3. numero di replicazioni o ripetizioni del disturbo r (valore compreso tra 1 e 999).

Figura 3



Sulla base di tali valori, per ciascun anno t , vengono generate r matrici di input, in cui si perturba il $p\%$ di unità statistiche scelte casualmente.

La formula di perturbazione è la seguente:

$$\hat{x}_{ijt} = x_{ijt} + \lambda [N(0, \sigma_{jt})]$$

dove:

\hat{x}_{ijt} è il valore perturbato dell'indicatore j , per l'unità i , al tempo t ;

x_{ijt} è il valore originario dell'indicatore j , per l'unità i , al tempo t ;

$[N(0, \sigma_{jt})]$ è una determinazione di una variabile casuale normale con media nulla e scostamento quadratico medio dell'indicatore j , al tempo t .

Quindi, su ciascuna delle r matrici perturbate, si calcolano gli indici compositi selezionati e le corrispondenti graduatorie. Tali graduatorie sono confrontate con quella originale (ottenuta dalla matrice di partenza), mediante il calcolo della media e dello scostamento quadratico medio delle differenze assolute di rango (shift).

Nella seconda pagina (figura 4), occorre selezionare gli indicatori elementari da elaborare (tramite una x) e definire la loro polarità (tramite + o -). Inoltre, è necessario specificare se i valori minimi e massimi da utilizzare per la normalizzazione degli indicatori elementari (goalpost) devono essere calcolati dal software o forniti dall'esterno (tramite y o n). Nel primo

caso, saranno considerati, per ciascun indicatore elementare, il minimo e il massimo tra tutte le unità statistiche, in tutti gli anni considerati.

Figura 4

```
Comando ==>
  Lista degli indicatori base e loro polarità
=====
Dati per REGIONE secondo la variabile: REGIONE
Anni disponibili: da 2006 a 2010
-----
X IND_1_TOT: ..... +
X IND_2_TOT: ..... +
X IND_3_TOT: ..... +
X IND_4_TOT: ..... -
X IND_6_TOT: ..... -
X IND_7_TOT: ..... -
X IND_8_TOT: ..... -
X IND_9_TOT: ..... -
-----
Calcolare dal file input i valori min e max degli indicatori ? (y/n) Y
-----
Selezionare almeno un indicatore base (x) e la sua polarità (-/+ ) e
premere <ENTER> oppure <F3> per continuare
```

La terza pagina (figura 5) è dedicata alla definizione dei pesi degli indicatori. Ciascun valore deve essere compreso tra 0,001 e 100. Per default, tutti gli indicatori hanno lo stesso peso, pari a 1.

Inoltre, occorre specificare se si desidera ponderare le unità statistiche con i pesi contenuti nel file di input (tramite y o n). Per default, i pesi della colonna 4 del file di input sono ignorati e tutte le unità statistiche hanno lo stesso peso, pari a 1.

Nella quarta pagina (figura 6) sono riportati i valori minimi e massimi calcolati dal software o già usati in precedenza che possono essere confermati o modificati dall'utente. Quindi, occorre specificare l'anno di riferimento per la normalizzazione (tale valore ha senso soltanto se, nel file di input, ci sono più anni da analizzare).

Infine (figura 7), è necessario selezionare gli anni per i quali effettuare l'elaborazione. Nel caso in cui il file di input contenga più anni, infatti, l'utente potrebbe essere interessato solo ad alcuni di essi.

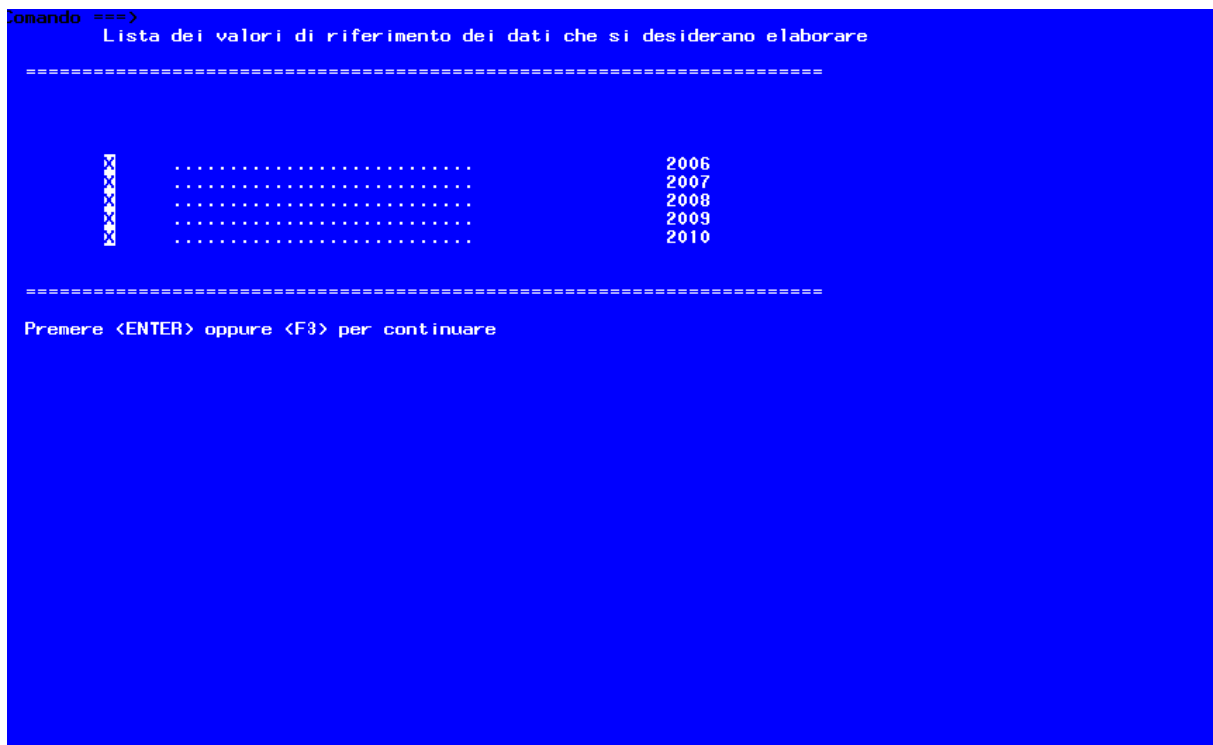
Figura 5

```
Domando ==>
  Lista degli indicatori base selezionati e relativo peso
=====
  Pesì (valore di default = 1):
-----
  IND_1_TOT: ..... 00001
  IND_2_TOT: ..... 00001
  IND_3_TOT: ..... 00001
  IND_4_TOT: ..... 00001
  IND_6_TOT: ..... 00001
  IND_7_TOT: ..... 00001
  IND_8_TOT: ..... 00001
  IND_9_TOT: ..... 00001
-----
  Si desidera ponderare con pesì >= 1 le unità statistiche del file di input? (y/n) 
  (n = dati non pesati; y = pesì da file in: esempio1.xls)
=====
  Il peso di ciascun indicatore può variare in un intervallo da 0.001 a 100.0 (1 valore 0 non è ammesso).
  Premere <ENTER> oppure <F3> per continuare
```

Figura 6

```
Domando ==>
  Lista degli indicatori base selezionati e relativo intervallo
=====
  Valori minimi e massimi usati già in precedenza:
-----
  IND_1_TOT: ..... 43.679385 75.804192
  IND_2_TOT: ..... 3.692561 36.048073
  IND_3_TOT: ..... 10.224992 37.774694
  IND_4_TOT: ..... 6.438937 36.518396
  IND_6_TOT: ..... 8.722067 29.78664
  IND_7_TOT: ..... 9.6031955826 26.807595864
  IND_8_TOT: ..... 4.4478527607 29.149991996
  IND_9_TOT: ..... 57.669446104 88.320163792
-----
  Anno del totale di riferimento per la standardizzazione: 2006
=====
  Premere <ENTER> oppure <F3> per continuare
```

Figura 7



Premendo F3 alla fine di ogni pagina, è possibile passare alla pagina successiva e, quindi, avviare l'elaborazione dei dati che può richiedere anche diversi minuti a seconda delle funzioni che sono state richieste.

7. Gli output

Durante l'esecuzione del software, viene creata automaticamente una directory di output con lo stesso nome del file di input (se questa esiste, viene sovrascritta). Tale directory è contenuta nella directory DATI\OUTPUT o in qualsiasi altra directory scelta dall'utente.

L'albero della directory di output è il seguente:

- DATI: dove sono salvati i data-set SAS di output (graduatorie, serie storiche, ecc.);
- WORD: dove sono salvati tutti i report in rich text format (.rtf);
- HTML: dove sono salvati tutti i report in html (.html).

Per facilitare la navigazione tra i report, un indice dei risultati viene creato automaticamente nella directory di output con un file ipertestuale in html (sommario.html).

L'indice dei risultati

Aperto il file sommario.html compare l'indice dei risultati (figure 8-11). A partire dall'indice, cliccando sulle voci evidenziate in blu, è possibile visualizzare prospetti riepilogativi, grafici e tabelle.

Figura 8

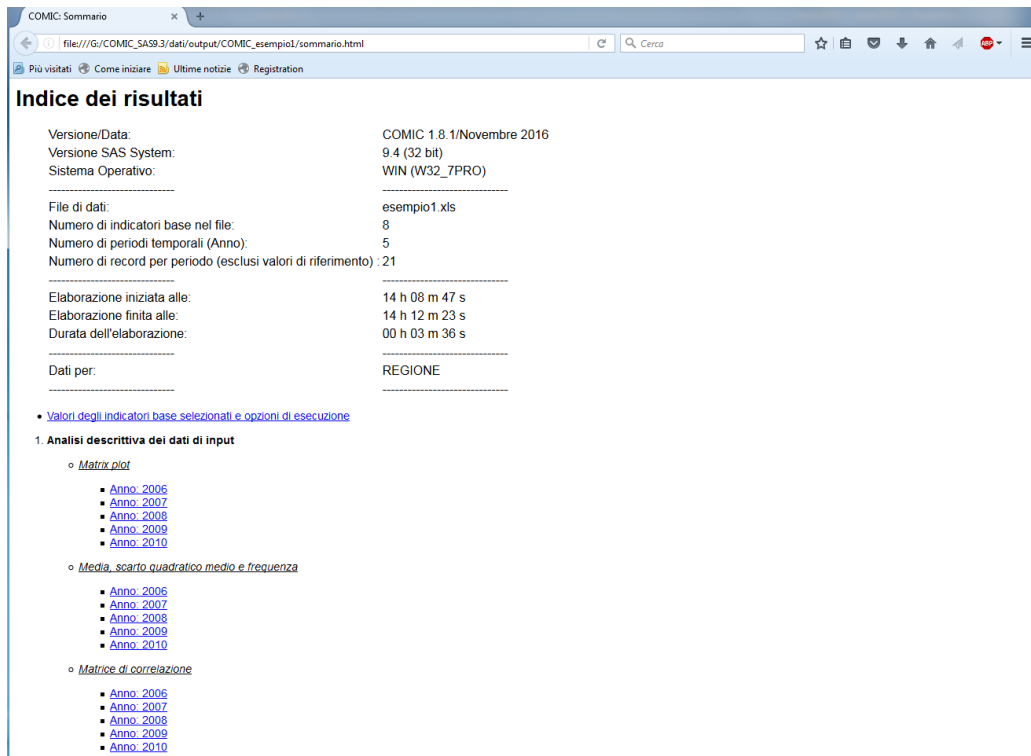


Figura 9

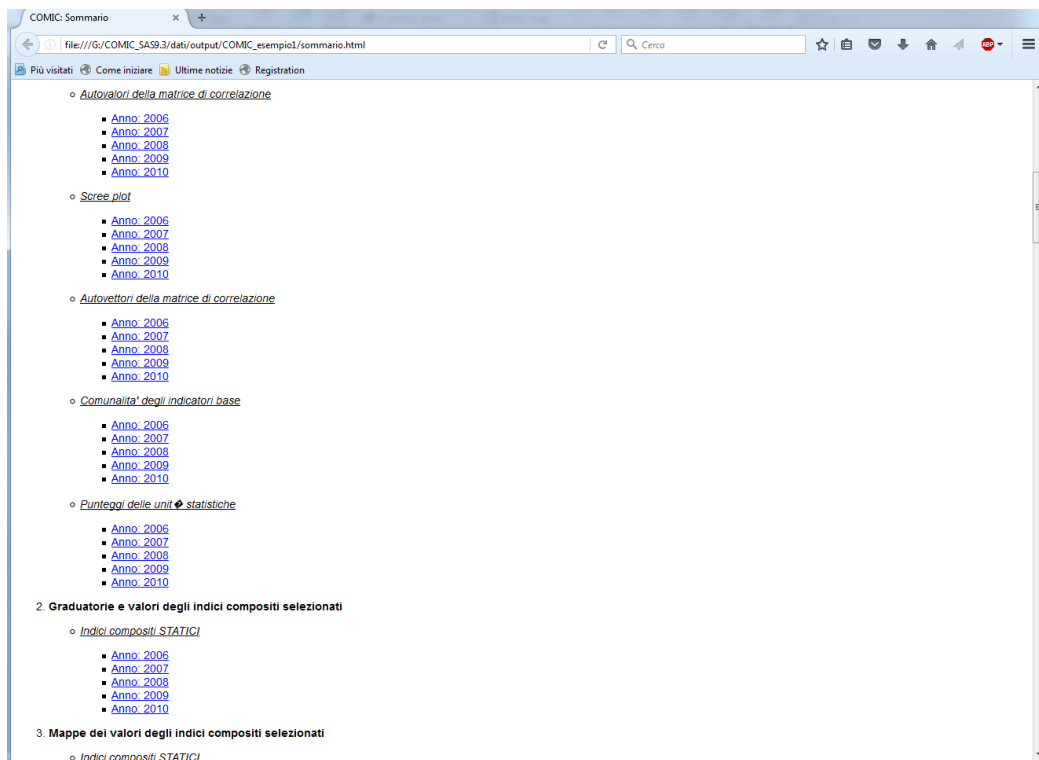


Figura 10

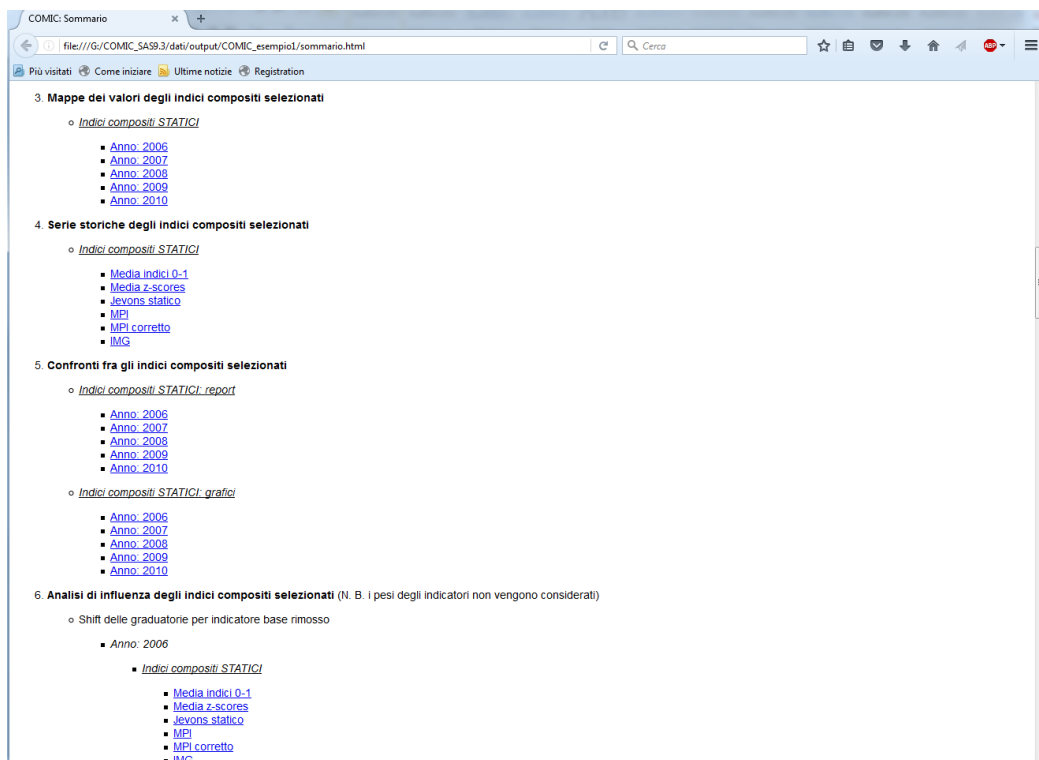
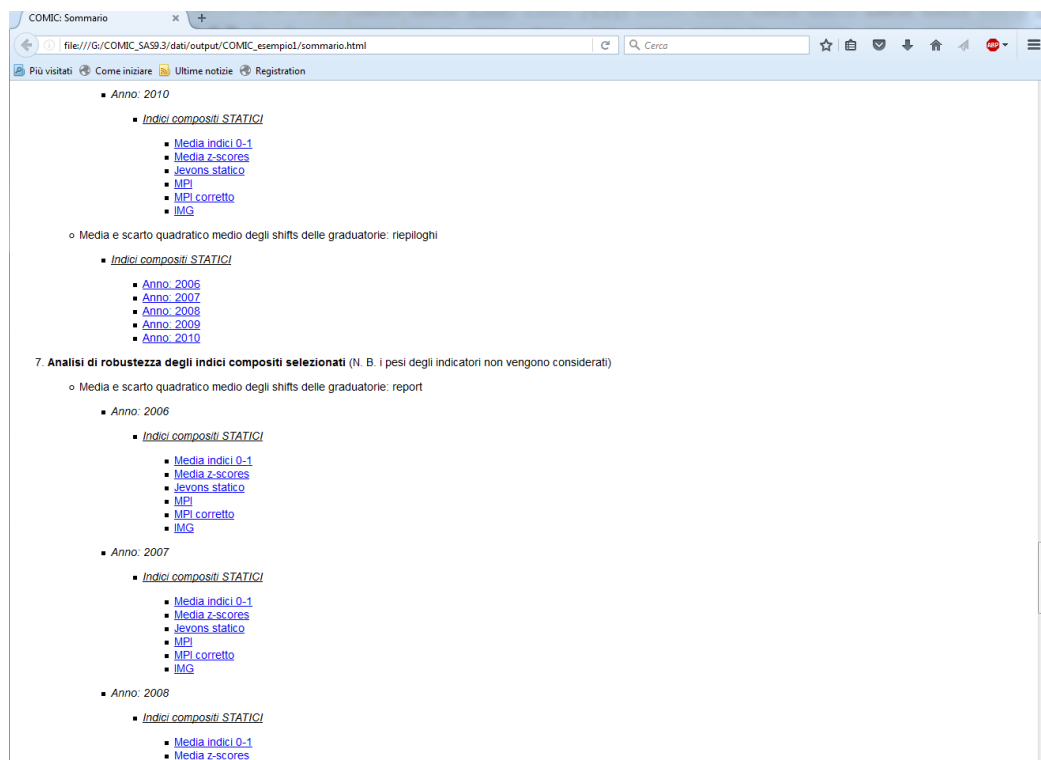


Figura 11



Grafici e tabelle

Le informazioni visualizzabili a partire dall'indice dei risultati sono le seguenti:

- Valori degli indicatori base selezionati e opzioni di esecuzione;
- Analisi descrittiva dei dati di input (per ogni anno):
 - Matrix plot degli indicatori selezionati;
 - Media, scarto quadratico medio e frequenza degli indicatori selezionati;
 - Matrici di correlazione degli indicatori selezionati;
 - Autovalori della matrice di correlazione degli indicatori selezionati;
 - Scree plot dell'Analisi in Componenti Principali (ACP) degli indicatori selezionati;
 - Autovettori della matrice di correlazione degli indicatori selezionati;
 - Comunalita' degli indicatori selezionati;
 - Punteggi delle unità statistiche ottenuti dall'ACP (Componenti Principali);
- Graduatorie e valori degli indici composti selezionati (per ogni anno);
- Mappe dei valori degli indici composti selezionati (per ogni anno);
- Serie storiche degli indici composti selezionati (per ciascun indice sintetico selezionato);
- Grafici velocità - accelerazione degli indici composti (per ogni anno e ciascun indice sintetico dinamico selezionato);
- Confronti fra gli indici composti selezionati: report e grafici (per ogni anno);

- *Analisi di influenza degli indici compositi selezionati* (N. B. i pesi degli indicatori elementari non vengono considerati);
 - *Shift delle graduatorie per indicatore base rimosso* (per ogni anno e ciascun indice sintetico selezionato);
 - *Media e scarto quadratico medio degli shifts delle graduatorie: report* (per ogni anno e ciascun indice sintetico selezionato);
 - *Media e scarto quadratico medio degli shifts delle graduatorie: grafici* (per ogni anno e ciascun indice sintetico selezionato);
 - *Media e scarto quadratico medio degli shifts delle graduatorie: riepiloghi* (per ogni anno);
- *Analisi di robustezza degli indici compositi selezionati* (N. B. i pesi degli indicatori non vengono considerati)
 - *Media e scarto quadratico medio degli shifts delle graduatorie: report* (per ogni anno e ciascun indice sintetico selezionato);
 - *Media e scarto quadratico medio degli shifts delle graduatorie: grafici* (per ogni anno e ciascun indice sintetico selezionato);
 - *Media e scarto quadratico medio degli shifts delle graduatorie: riepiloghi* (per ogni anno).

Nelle pagine seguenti sono illustrati alcuni esempi dei risultati.

In figura 12 è mostrata la pagina relativa a “*Valori degli indicatori base selezionati e opzioni di esecuzione*”. La figura 13 rappresenta il “*Matrix plot degli indicatori selezionati*”, dove ciascun indicatore selezionato è messo in relazione con gli altri, al fine di ‘visualizzare’ le relazioni esistenti. Nella figura 14 è riportata la “*Matrice di correlazione*” degli indicatori selezionati; mentre in figura 15 è illustrato lo “*Scree plot*” dell’Analisi in Componenti Principali, da cui è possibile desumere la presenza di eventuali dimensioni latenti.

La figura 16 contiene la tabella riguardante “*Graduatorie e valori degli indici compositi selezionati*”; tale tabella include anche il totale di riferimento (in tal caso, la media nazionale). I valori degli indici compositi sono rappresentati graficamente mediante le “*Mappe dei valori degli indici compositi selezionati*”, illustrate in figura 17.

Nella figura 18 è riportata una “*Serie storiche degli indici compositi selezionati*”; mentre la figura 19 mostra un esempio di “*Grafici velocità - accelerazione degli indici compositi*”.

Infine, le figure 20-21 riguardano i “*Confronti fra gli indici compositi selezionati: report e grafici*”; le figure 22-23 l’Analisi di Influenza e le figure 24-25 l’Analisi di Robustezza.

Figura 12

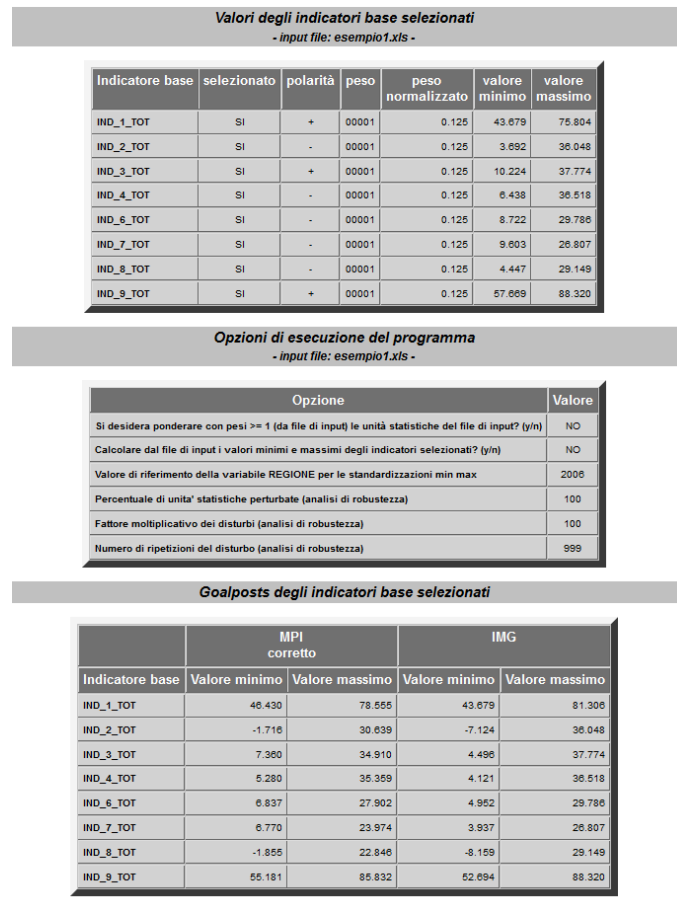


Figura 13

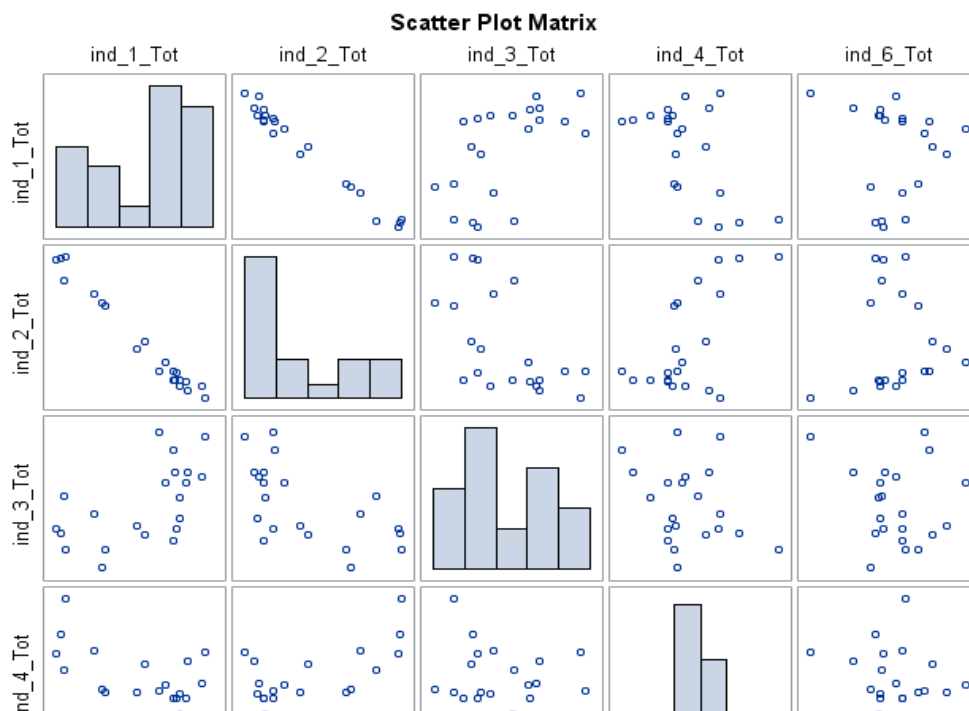


Figura 14

Matrice di correlazione dei dati in: esempio1.xls
- Anno: 2006 -

Indicatore base	ind_1_Tot	ind_2_Tot	ind_3_Tot	ind_4_Tot	ind_6_Tot	ind_7_Tot	ind_8_Tot	ind_9_Tot
ind_1_Tot	1.000	-0.988	0.581	-0.571	-0.094	0.300	-0.915	0.608
ind_2_Tot	-0.988	1.000	-0.594	0.625	0.079	-0.265	0.933	-0.610
ind_3_Tot	0.581	-0.594	1.000	-0.299	-0.160	0.204	-0.597	0.348
ind_4_Tot	-0.571	0.625	-0.299	1.000	-0.225	0.035	0.594	-0.597
ind_6_Tot	-0.094	0.079	-0.160	-0.225	1.000	0.075	0.212	0.490
ind_7_Tot	0.300	-0.265	0.204	0.035	0.075	1.000	-0.116	0.258
ind_8_Tot	-0.915	0.933	-0.597	0.594	0.212	-0.116	1.000	-0.432
ind_9_Tot	0.608	-0.610	0.348	-0.597	0.490	0.258	-0.432	1.000

Figura 15

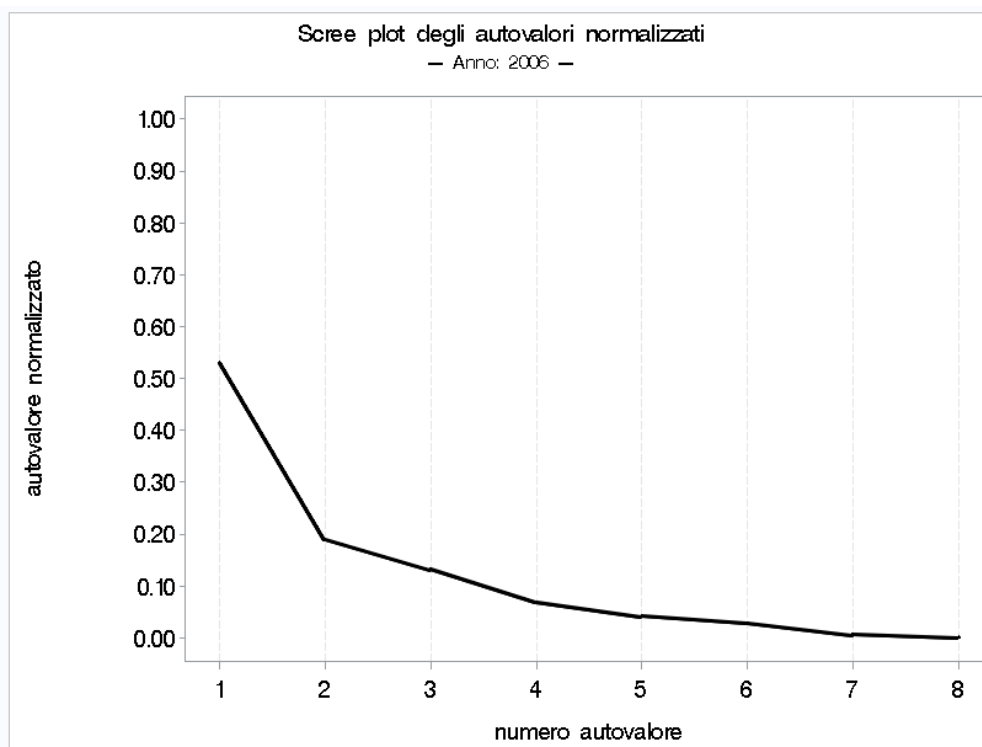


Figura 16

Indicatori compositi e graduatorie
- Anno: 2006 -

Regione	Media indici 0-1		Media z-scores		Jevons statico		MPI		MPI corretto		IMG	
	Valore	Rango	Valore	Rango	Valore	Rango	Valore	Rango	Valore	Rango	Valore	Rango
Piemonte	0.776	1.0	0.859	1.0	133.233	2.0	108.455	1.0	111.600	1.0	131.523	1.0
Valle d'Aosta/Vallée d'Aoste	0.716	4.0	0.617	4.0	130.919	4.0	105.711	4.0	106.962	4.0	116.421	4.0
Liguria	0.705	6.0	0.509	7.0	113.801	9.0	104.512	7.0	106.097	6.0	116.263	5.0
Lombardia	0.754	2.0	0.746	2.0	131.541	3.0	107.315	2.0	110.479	2.0	127.671	2.0
Bolzano/Bozen	0.666	8.0	0.438	8.0	135.921	1.0	101.995	9.0	100.829	11.0	83.450	13.0
Trento	0.663	9.0	0.376	9.0	121.716	8.0	103.528	8.0	104.543	8.0	111.539	8.0
Veneto	0.738	3.0	0.704	3.0	129.334	5.0	106.980	3.0	109.438	3.0	125.306	3.0
Friuli-Venezia Giulia	0.626	10.0	0.198	10.0	108.995	11.0	101.696	10.0	102.013	9.0	104.613	9.0
Emilia-Romagna	0.703	7.0	0.524	6.0	122.431	7.0	104.898	5.0	106.115	5.0	114.845	6.0
Toscana	0.618	11.0	0.178	11.0	109.003	10.0	101.456	11.0	101.232	10.0	101.949	10.0
Umbria	0.568	12.0	-0.098	12.0	96.948	12.0	97.398	13.0	95.070	14.0	58.637	20.0
Marche	0.709	5.0	0.551	5.0	123.492	6.0	104.669	6.0	105.739	7.0	113.704	7.0
Lazio	0.556	13.0	-0.144	13.0	96.297	13.0	97.960	12.0	98.087	12.0	93.516	11.0
Abruzzo	0.529	14.0	-0.238	14.0	94.232	14.0	97.273	14.0	96.227	13.0	89.170	12.0
Molise	0.424	17.0	-0.636	17.0	80.943	17.0	93.425	17.0	90.138	17.0	74.591	15.0
Campania	0.399	19.0	-0.706	19.0	80.974	16.0	91.745	19.0	87.372	19.0	63.146	18.0
Puglia	0.446	16.0	-0.505	16.0	86.856	15.0	94.001	16.0	90.848	15.5	72.924	16.0
Basilicata	0.408	18.0	-0.697	18.0	80.532	19.0	92.886	18.0	89.483	18.0	72.349	17.0
Calabria	0.253	21.0	-1.362	21.0	67.091	21.0	85.479	21.0	78.488	21.0	29.216	21.0
Sicilia	0.361	20.0	-0.856	20.0	76.897	20.0	90.614	20.0	85.691	20.0	59.103	19.0
Sardegna	0.451	15.0	-0.457	15.0	80.814	18.0	94.808	15.0	90.848	15.5	75.286	14.0
ITALIA	0.577		0.005		100.000		99.940		100.000		100.000	

Figura 17



Figura 18

Serie storica dell'indicatore composito
- Media indici 0-1 -

Regione	2006		2007		2008		2009		2010	
	Valore	Rango	Valore	Rango	Valore	Rango	Valore	Rango	Valore	Rango
Piemonte	0.776	1.0	0.731	3.0	0.780	3.0	0.715	4.0	0.694	5.0
Valle d'Aosta/Vallée d'Aoste	0.716	4.0	0.680	8.0	0.747	5.0	0.680	8.0	0.771	1.0
Liguria	0.705	6.0	0.694	7.0	0.647	11.0	0.636	10.0	0.693	6.0
Lombardia	0.754	2.0	0.778	1.0	0.799	1.0	0.747	3.0	0.722	2.0
Bolzano/Bozen	0.666	8.0	0.652	9.0	0.610	12.0	0.644	9.0	0.582	11.0
Trento	0.663	9.0	0.695	6.0	0.683	8.0	0.692	7.0	0.636	8.0
Veneto	0.738	3.0	0.759	2.0	0.792	2.0	0.750	2.0	0.719	3.0
Friuli-Venezia Giulia	0.626	10.0	0.696	5.0	0.754	4.0	0.755	1.0	0.699	4.0
Emilia-Romagna	0.703	7.0	0.708	4.0	0.715	6.0	0.705	5.0	0.637	7.0
Toscana	0.618	11.0	0.648	10.0	0.710	7.0	0.702	6.0	0.605	10.0
Umbria	0.568	12.0	0.601	12.0	0.597	13.0	0.555	13.0	0.529	13.0
Marche	0.709	5.0	0.647	11.0	0.653	10.0	0.628	11.0	0.627	9.0
Lazio	0.556	13.0	0.586	13.0	0.592	14.0	0.576	12.0	0.569	12.0
Abruzzo	0.529	14.0	0.541	14.0	0.667	9.0	0.526	14.0	0.479	14.0
Molise	0.424	17.0	0.516	15.5	0.507	15.5	0.437	17.0	0.455	15.0
Campania	0.399	19.0	0.374	18.0	0.405	18.0	0.374	18.0	0.324	19.0
Puglia	0.446	16.0	0.457	17.0	0.464	17.0	0.441	16.0	0.378	16.0
Basilicata	0.408	18.0	0.357	19.0	0.359	19.0	0.296	19.0	0.343	18.0
Calabria	0.253	21.0	0.241	21.0	0.265	21.0	0.219	21.0	0.189	21.0
Sicilia	0.361	20.0	0.352	20.0	0.319	20.0	0.271	20.0	0.241	20.0
Sardegna	0.451	15.0	0.516	15.5	0.507	15.5	0.459	15.0	0.375	17.0
ITALIA	0.577		0.588		0.606		0.574		0.539	

Figura 19

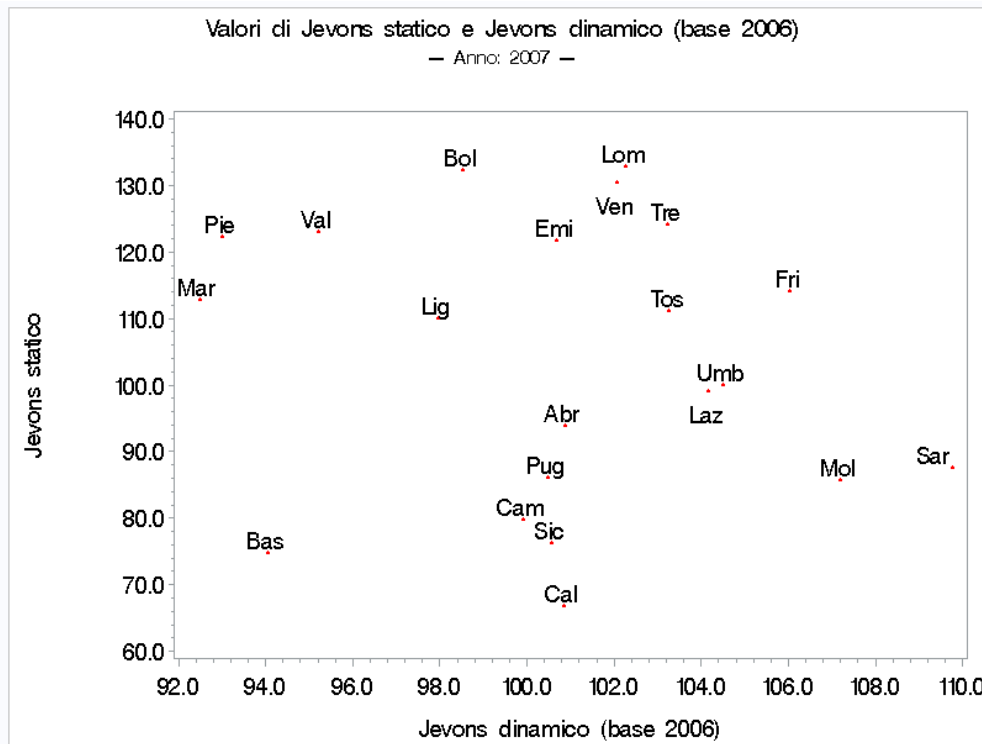


Figura 20

Matrice di correlazione di Spearman dei metodi selezionati
- Anno: 2006 -

Indicatore composito	Media indici 0-1	Media z-scores	Jevons statico	MPI	MPI corretto	IMG
Media indici 0-1	1.000	0.999	0.943	0.994	0.983	0.926
Media z-scores	0.999	1.000	0.945	0.996	0.984	0.925
Jevons statico	0.943	0.945	1.000	0.934	0.905	0.822
MPI	0.994	0.996	0.934	1.000	0.993	0.944
MPI corretto	0.983	0.984	0.905	0.993	1.000	0.965
IMG	0.926	0.925	0.822	0.944	0.965	1.000

Matrice delle differenze medie assolute di rango dei metodi selezionati
- Anno: 2006 -

Indicatore composito	Media indici 0-1	Media z-scores	Jevons statico	MPI	MPI corretto	IMG
Media indici 0-1	0.000	0.095	1.238	0.381	0.714	1.429
Media z-scores	0.095	0.000	1.238	0.286	0.714	1.429
Jevons statico	1.238	1.238	0.000	1.333	1.571	2.286
MPI	0.381	0.286	1.333	0.000	0.429	1.238
MPI corretto	0.714	0.714	1.571	0.429	0.000	0.905
IMG	1.429	1.429	2.286	1.238	0.905	0.000

Figura 21

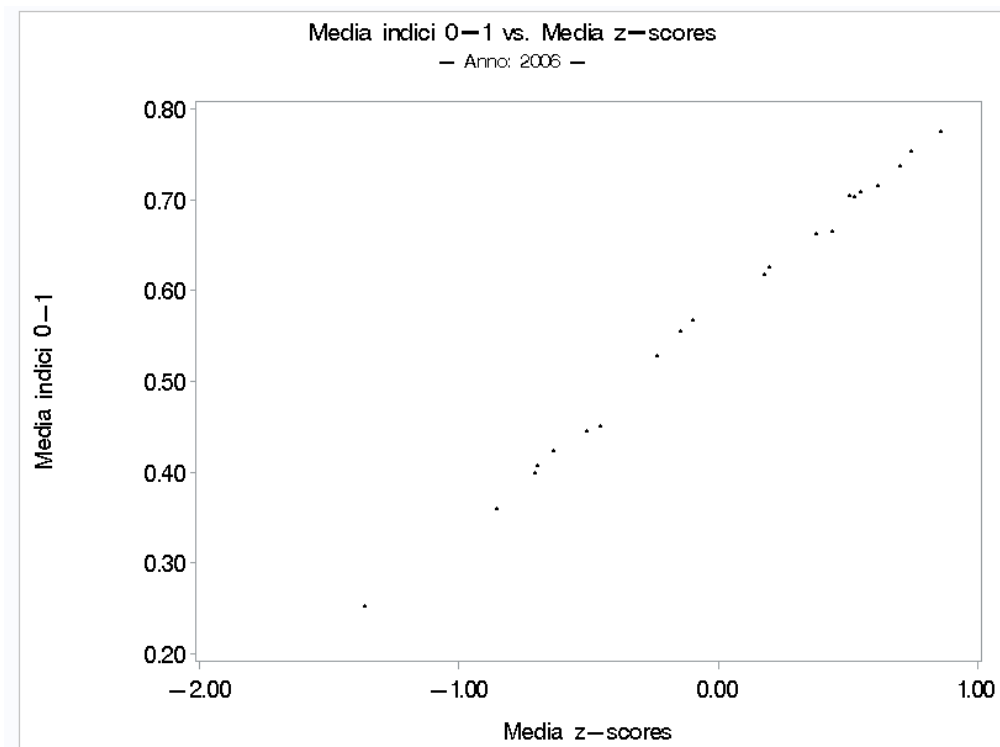


Figura 22

Media e scarto quadratico medio degli shifts delle graduatorie per indicatore base rimosso
 - Tipo di indicatore composito: Media indici 0-1 - Anno: 2006

Indicatore base rimosso	Media	Scarto quadratico medio
ind_1_tot	0.667	0.713
ind_2_tot	0.667	0.642
ind_3_tot	1.048	1.045
ind_4_tot	0.857	0.774
ind_6_tot	0.857	0.953
ind_7_tot	1.381	1.371
ind_8_tot	0.238	0.590
ind_9_tot	0.857	1.037
Media	0.821	0.891
S.q.m.	0.307	0.244

Figura 23

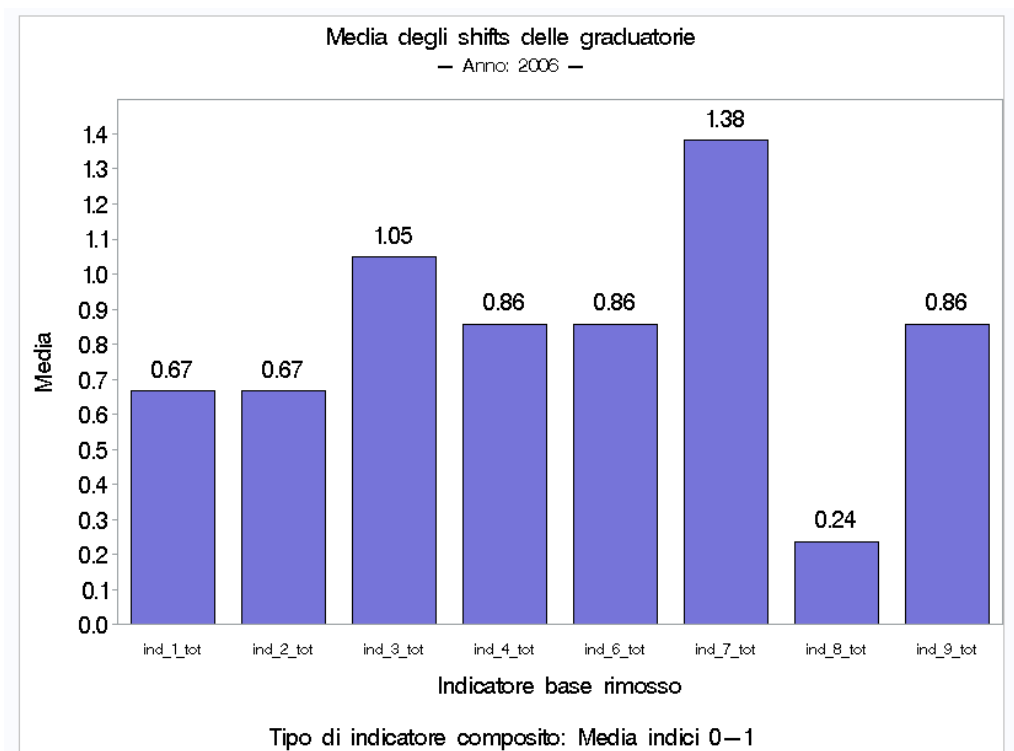
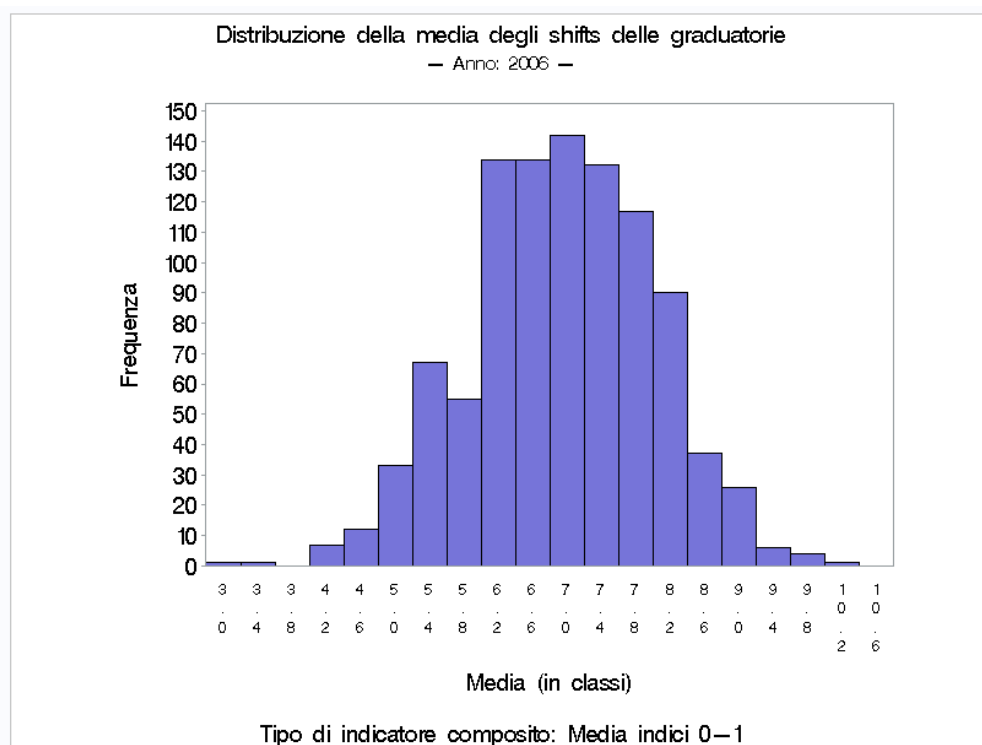


Figura 24

Media e scarto quadratico medio degli shifts delle graduatorie per ogni perturbazione
 - Tipo di indicatore composito: Media indici 0-1 - Anno: 2006

Perturbazione	Media	Scarto quadratico medio
test001	6.952	5.447
test002	6.952	4.306
test003	5.381	4.889
test004	6.810	3.781
test005	5.571	5.181
test006	6.952	4.554
test007	6.190	5.172
test008	8.857	5.642
test009	6.524	4.579
test010	7.905	4.242
test011	7.571	5.549
test012	5.762	3.819
test013	7.714	5.191
test014	8.095	5.797
test015	6.571	5.473
test016	7.619	5.131

Figura 25



8. Sviluppi futuri

I possibili sviluppi di COMIC riguardano i seguenti aspetti:

- implementazione di procedure SAS per migliorare l'analisi esplorativa dei dati di input;
- possibilità di creare delle mappe geografiche “Mondo” ed “Europa”;
- versione alternativa in linguaggio R.

Riferimenti bibliografici

Aureli Cutillo E. (1996), *Lezioni di statistica sociale. Parte seconda, sintesi e graduatorie*, CISU, Roma.

Delvecchio F. (1995), *Scale di misura e indicatori sociali*, Cacucci editore, Bari.

Mazziotta M., Pareto A. (2011), *Un indice sintetico non compensativo per la misura della dotazione infrastrutturale: un'applicazione in ambito sanitario*, *Rivista di Statistica Ufficiale*, 1/2011, pp. 63-79.

Mazziotta, M., Pareto, A. (2016), *On a Generalized Non-compensatory Composite Index for Measuring Socio-economic Phenomena*, *Social Indicators Research*, 127, pp. 983-1003.

OECD (2008), *Handbook on Constructing Composite Indicators. Methodology and user guide*, OECD Publications, Paris.

Tarantola, S. (2008), *European Innovation Scoreboard: strategies to measure country progress over time*, JRC Scientific and Technical Reports, EUR 23526 EN, Luxembourg.

UNDP (2010), *Human Development Report 2010*, Palgrave Macmillan, New York.