

Analisi fattoriale ed elaborazioni veloci

Ulisse Di Corpo

Sintropia-DS può gestire archivi di qualsiasi dimensione (in questa versione il programma è stato limitato ad archivi di 500.000 schede). Come già detto Sintropia-DS organizza i dati e gli archivi in modo da ottimizzare l'analisi statistica dei dati. Nonostante queste ottimizzazioni al crescere del numero delle schede crescono anche i tempi di attesa per le elaborazioni. Quando le schede sono decine di migliaia o centinaia di migliaia e le variabili molte i tempi di attesa possono allungarsi eccessivamente, scoraggiando quindi l'uso continuo e interattivo delle analisi statistiche. Per ovviare a questi problemi è stata inserita la sezione "Elaborazioni Veloci" che consente, indipendentemente dalle dimensioni degli archivi, di realizzare elaborazioni "istantanee".

Le "elaborazioni veloci" utilizzano griglie delle variabili dicotomiche e quantitative che in genere corrispondono alla copia esatta delle griglie utilizzate nella sezione base delle elaborazioni; a queste griglie si accede grazie al primo tasto "Griglie – definizione delle variabili". Inoltre, questa sezione per consentire elaborazioni istantanee richiede una diversa organizzazione dei dati, questo è il motivo che ha portato all'inserimento del tasto "Prepara le elaborazioni veloci".

Mentre le griglie rimangono, in genere, stabili nel tempo, la preparazione delle elaborazioni veloci deve essere utilizzata ogni qual volta il file dati cambia e si vogliono aggiornare le elaborazioni.

Una volta che gli archivi sono stati preparati (in genere si tratta di una operazione che richiede pochi secondi), è possibile:

1. **selezionare le schede.** Le selezioni effettuate in questa sezione possono essere utilizzate in tutte le altre sezioni del programma.
2. effettuare **distribuzioni semplici.** Le finestre delle analisi dei dati sono rimaste praticamente identiche a quelle precedenti. L'unica differenza è l'assenza del tasto "Elabora". Essendo la procedura di calcolo istantanea il programma ricalcola i risultati ogni volta che si chiede di Visualizzare i risultati.
3. **tabelle doppie e connessioni.** Per chiarezza espositiva le tabelle doppie e le connessioni fanno



adesso parte della stessa finestra. In questo modo l'utilizzatore viene portato a considerare queste due forme di restituzione dei dati analoghe.

4. **reports.** E' possibile predisporre delle interrogazioni che vengono attivate selezionando il nome dell'interrogazione ed è possibile visualizzare il risultato in base a tabelle le cui colonne e righe possono essere il frutto di elaborazioni operate su più variabili. In questo modo è possibile la restituzione istantanea di combinazioni complesse di interrogazioni e di analisi dei dati.
5. **analisi fattoriali e tipologie.** L'analisi fattoriale mantiene il tasto "calcola i fattori", in quanto i tempi di elaborazione sono principalmente dovuti all'analisi fattoriale, piuttosto che ai tempi di lettura dei dati, e in quanto dopo il calcolo dei fattori è necessario ruotare gli assi fattoriali.
6. **liste.** Anche la produzione delle liste è stata ottimizzata. I tempi di attesa rimangono però proporzionali alle dimensioni dell'archivio in quanto questa opzione continua ad accedere ai dati in modo "tradizionale".

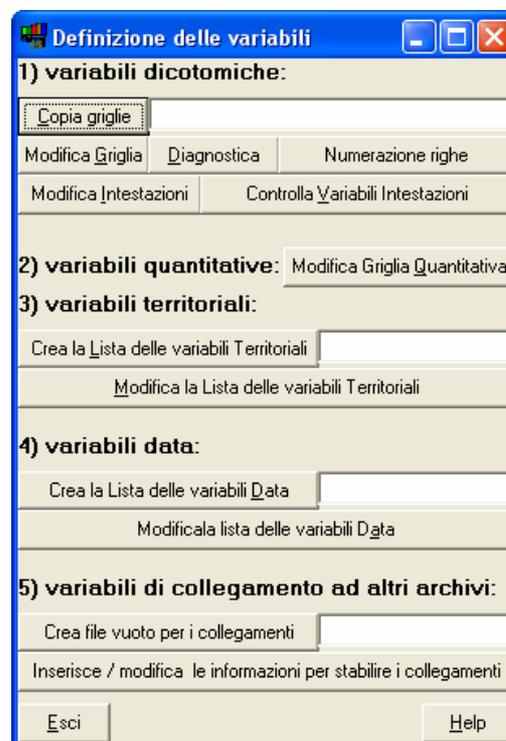
Altre caratteristiche delle elaborazioni veloci sono:

- collegamento con altri archivi. E' possibile selezionare le schede in base a condizioni inserite su archivi collegati (ad esempio se si gestisce un archivio fatture, gli archivi collegati potrebbero essere l'archivio clienti e l'archivio rappresentanti).
- variabili aggiuntive. E' possibile creare variabili dicotomiche aggiuntive come risultato di elaborazioni effettuate sulle variabili dicotomiche di partenza. Le variabili aggiuntive possono essere di due tipi: similarità a profili fattoriali (o profili di qualsiasi altra natura), risultato di una o più condizioni inserite sulle variabili dicotomiche di partenza.

Griglia, definizione delle variabili

Questa sezione viene in genere utilizzata da colui che realizza l'applicazione e serve per impostare le elaborazioni veloci. Le elaborazioni veloci richiedono la definizione delle variabili dicotomiche, ma possono anche richiedere la definizione delle variabili quantitative, territoriali, data e collegamenti:

1. **variabili dicotomiche.** In questa sezione spesso si utilizza una semplice copia della griglia delle elaborazioni normali. Può risultare utile modificare la griglia delle variabili dicotomiche in modo da rendere più agevole l'utilizzo del programma (se modificate la griglia ricordatevi sempre di utilizzare il tasto "Diagnostica" per controllare se incidentalmente non aveste inserito degli errori). Il file griglia è `grl-dat.txt`, mentre il file con le intestazioni delle tabelle è `tin-dat.txt`. E' importante tenere allineato il file griglia con il file intestazioni, per questo motivo è stato inserito il tasto "Controlla Variabili / Intestazioni".
2. **variabili quantitative.** La griglia delle variabili quantitative viene copiata automaticamente (quando si effettua la copia della griglia delle variabili dicotomiche) dalla griglia quantitativa utilizzata per le elaborazioni normali. Consigliamo di modificare questa griglia in modo da abbinare alle variabili quantitative sigle migliori più leggibili di quelle che vengono proposte in automatico dal



programma.

3. **variabili territoriali.** Nella scheda possono essere presenti più variabili territoriali (campi che fanno riferimento alla lista dei comuni ISTAT). In questa sezione si crea la lista delle variabili territoriali, ed è possibile intervenire modificando i nomi con i quali queste variabili vengono identificate.
4. **variabili data.** Nella scheda possono essere presenti più variabili data. In questa sezione si crea la lista delle variabili data, ed è possibile intervenire modificando i nomi con i quali queste variabili vengono identificate.
5. **variabili di collegamento ad altri archivi.**

Prepara le elaborazioni veloci

In questa sezione vengono preparati i file che consentono di realizzare le elaborazioni veloci. A seconda delle griglie che sono state attivate nella sezione precedente saranno attivi uno o più bottoni. Per i bottoni “Prepara la variabile Quantitativa”, “Prepara la variabile Territoriale”, “Prepara la variabile Data”, è necessario



selezionare la variabile da preparare. In questa versione del programma si è infatti scelto di limitare le elaborazioni veloci solo ad una sola variabile territoriale / data / quantitativa, in modo da ridurre le opzioni di selezione nelle maschere di interrogazione dell'archivio. Selezionate le variabili da preparare e premete in sequenza i tasti “Prepara” che risultano essere attivi.

Quando esistono variabili aggiuntive (profili fattoriali o variabili che risultano da più condizioni inserite su altre variabili dicotomiche) è necessario entrare nella finestra “Variabili aggiuntive” e premere il tasto “Crea/Aggiorna le variabili”.

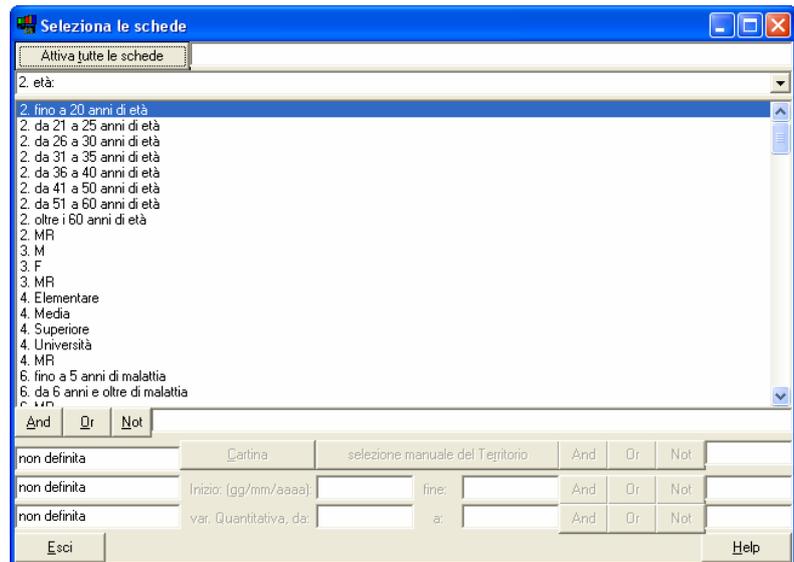
Seleziona le schede

Quando si utilizzano le opzioni di preparazione delle elaborazioni veloci il programma azzerava la selezione delle schede. Una volta effettuata una selezione, il programma tiene memoria delle schede selezionate. Quando effettuate una nuova selezione iniziate premendo il primo tasto “Attiva tutte le schede”.

La selezione avviene operando sulle variabili dicotomiche, la variabile quantitativa selezionata nella finestra “Prepara le elaborazioni veloci”, la variabile data e la variabile territorio. Il nome della variabile quantitativa, territorio e data attualmente attive vengono mostrate nei riquadri a sinistra della finestra.

Per effettuare le selezioni si opera come segue:

- Per le variabili dicotomiche si seleziona la variabile che interessa e si preme il tasto “And”, “Or” o “Not” a seconda della selezione che si intende eseguire. *Il tasto “And” seleziona (dalle schede già attive) solo le schede che presentano la caratteristica. Il tasto “Or” aggiunge alle schede già attive le schede che presentano la caratteristica. Il tasto “Not” disattiva tutte quelle schede che presentano la caratteristica selezionata.*
- Per la variabile quantitativa è necessario inserire l'intervallo entro il quale effettuare la selezione, quindi premere il tasto “And”, “Or” o “Not” a seconda della selezione che si vuole operare. *Come si vede dall'esempio nel momento in cui una variabile non è definita, le opzioni di selezione vengono disabilitate.*
- Per la variabile territoriale è necessario selezionare manualmente il territorio o per mezzo della cartina e a questo punto effettuare la selezione per mezzo dei tasti “And”, “Or” o “Not”.
- Per la variabile data è necessario inserire il periodo che interessa e attivare la ricerca. *Per la variabile data e la variabile quantitativa può essere inserito solo un valore dell'intervallo lasciando l'altro aperto. Ad esempio se si inserisce la data superiore, verranno selezionate tutte le schede fino a tale data*



Distribuzioni semplici, tabelle doppie, connessioni e analisi fattoriale

Le analisi dei dati utilizzano finestre simili a quelle delle elaborazioni normali, l'unica differenza è l'assenza del tasto “Elabora” e l'unificazione della finestra tabelle doppie con la finestra connessioni. Per il resto la logica di queste finestre, il modo di operare e di leggere i risultati rimane invariato.

Analisi fattoriale

I risultati dell'analisi delle connessioni, seppure molto più sintetici di quelli che si ottengono dallo studio delle tabelle doppie, spesso risultano troppo analitici: si osservano le singole relazioni, ma non come queste siano organizzate in strutture e gruppi. A tal fine si utilizza l'analisi fattoriale.

Le strutture di correlazione corrispondono a concetti generali, ad esempio aggressività, rendimento, ecc... In questo manuale le tecniche di analisi statistica sono state presentate partendo da quelle più analitiche, ad esempio le distribuzioni di frequenza, per poi passare alle tabelle doppie, all'analisi delle connessioni e adesso all'analisi fattoriale.

In genere quando si effettua un lavoro di ricerca si segue il percorso inverso. Si parte dall'analisi fattoriale, in modo da individuare i concetti generali, le strutture, che sottendono il fenomeno oggetto di indagine, e solo successivamente si passa alla ricerca di informazioni particolari, singole correlazioni, alcune tabelle particolarmente significative e distribuzioni semplici che fotografano il campione e le sue caratteristiche.



Per effettuare un'analisi fattoriale è necessario:

1. effettuare una opportuna selezione delle variabili;
2. calcolare i fattori e ruotare gli assi fattoriali;
3. visualizzare e interpretare le strutture fattoriali;
4. attribuire le tipologie (i pesi fattoriali) alle schede.

Selezione delle variabili

Nell'analisi fattoriale la selezione delle variabili è un momento particolarmente delicato. Infatti se si selezionano tutte le variabili emergeranno, probabilmente, strutture (fattori) generali, molto ovvie che schiacciano le strutture più fini che invece interessano ai fini dell'indagine. Diventa quindi importante effettuare una opportuna selezione delle variabili. In genere si seguono due strade:

1. si selezionano le variabili in base ad ipotesi teoriche o guidati dall'esperienza;
2. si selezionano le variabili in base alle correlazioni ottenute da una variabile chiave.

Per selezionare le variabili dovete premere il tasto seleziona. Verrà mostrata la finestra "Modalità selezionate", finestra simile a quella già esaminata per la selezione delle righe e delle colonne.

Selezione delle variabili in base ad ipotesi teoriche o guidati dall'esperienza

Quando si seguono ipotesi teoriche o si viene guidati dall'esperienza è bene evitare, nella selezione delle variabili/modalità, di inserire strutture che sono, banalmente, correlate tra loro.

Ad esempio, se esistono questionari con molte mancate risposte e nell'analisi fattoriale non vengono omesse le mancate risposte (MR) si osserverà, quasi certamente, un primo fattore composto da variabili "mancate risposte": in genere le mancate risposte correleranno molto tra di loro. Allo stesso modo è bene evitare di inserire variabili simmetriche, ad esempio se si inserisce la variabile "soddisfatto" è inutile inserire la variabile "insoddisfatto", ecc... in altre parole, quando esiste una polarità positiva e una polarità negativa è bene inserire una sola di queste due polarità (in genere quella positiva).

Inoltre, come si diceva prima, se si selezionano troppe variabili emergeranno strutture generali e in genere ovvie. Per far emergere strutture più interessanti si consiglia di utilizzare solo quelle variabili che possono essere utili per la costruzione dei fattori, delle strutture, che si stanno cercando.

Selezione delle variabili in base alle correlazioni ottenute da una variabile chiave

Spesso si utilizza l'analisi fattoriale per sintetizzare, raggruppare, le relazioni individuate per mezzo dell'analisi delle connessioni. La procedura è semplice: si effettua una prima analisi delle connessioni la si salva come file testo .txt. Se lo si ritiene necessario si tolgono da questo file le variabili che, in base alla nostra esperienza, non sono coerenti con l'oggetto dell'indagine. Ad esempio, se si prende l'archivio MonteOro (fornito con l'installazione di Sintropia-DS) e si calcolano le connessioni tra la variabile "Mi sento depresso" e le altre variabili del questionario si ottiene:

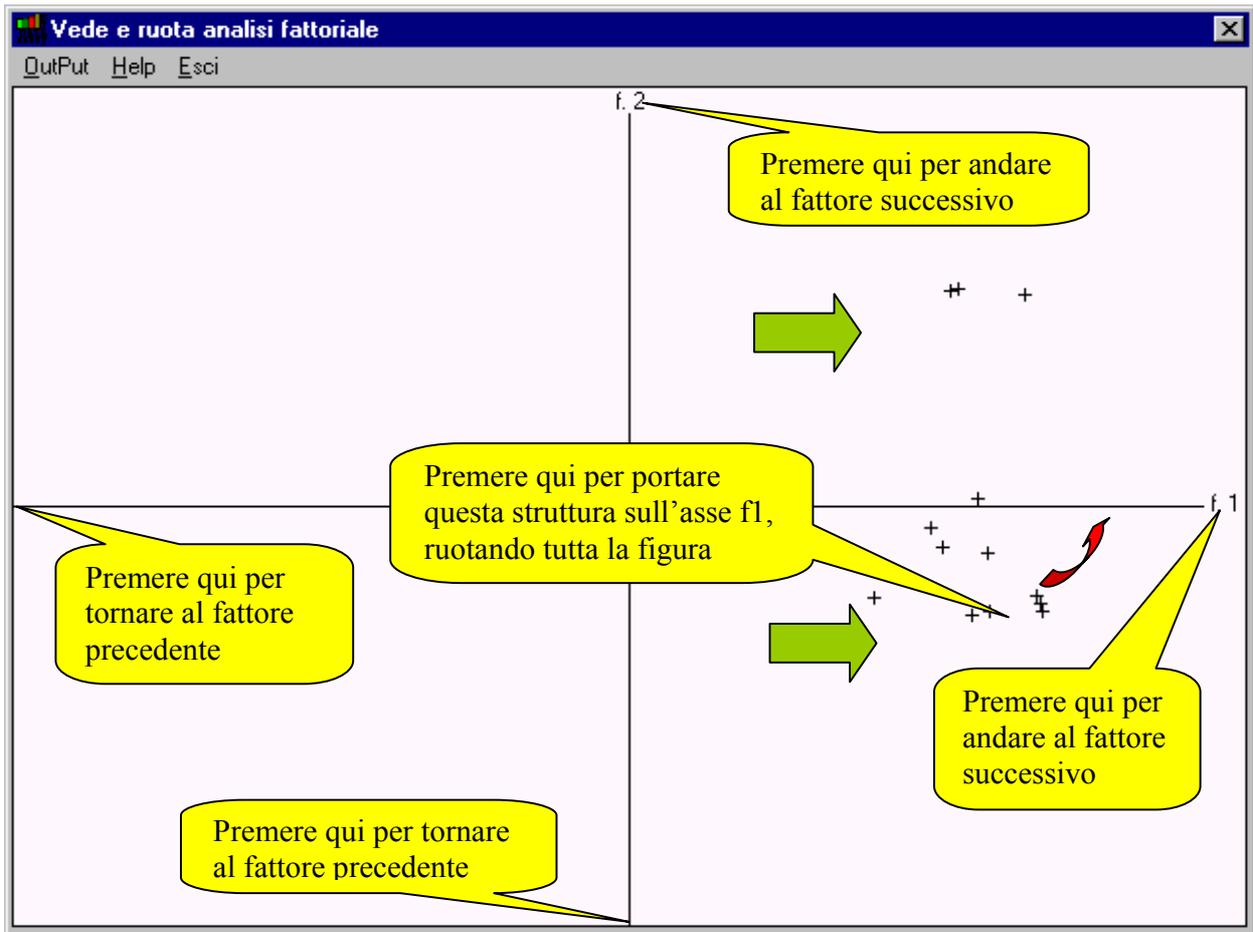
F/4 Mi sento depresso

ChiQ	rPhi			
974.00	1.000	(47)	(100.00%/ 0.00%)	F/4 Mi sento depresso
235.52	0.492	(46)	(80.73%/ 32.09%)	F/3 Mi sento angosciato
232.04	0.488	(50)	(64.55%/ 16.26%)	F/7 Ho poca fiducia nella vita
231.06	0.487	(49)	(70.52%/ 21.76%)	F/6 Mi sento solo
179.89	0.430	(51)	(66.47%/ 23.52%)	F/8 Non mi piace il mio corpo
93.66	0.310	(52)	(73.41%/ 42.86%)	F/9 Mi sento brutto
90.24	0.304	(123)	(48.94%/ 19.78%)	N/11 Mi sento molto impacciato
85.32	0.296	(45)	(36.22%/ 10.77%)	F/2 Ho poca fiducia in me stesso
59.72	0.248	(53)	(31.21%/ 10.77%)	F/10 Mi sento in grado di fare scelte per il mio futuro
56.90	0.242	(1)	(13.87%/ 0.88%)	D/1 E' principalmente una crisi economica
53.81	0.235	(120)	(69.36%/ 46.15%)	N/8 Mi piace molto essere il capo
49.87	0.226	(141)	(69.75%/ 47.47%)	Q/5 Mi sento impacciato/a con la sessualità

Prima di far leggere queste modalità all'analisi fattoriale (utilizzando il tasto Legge le connessioni che si trova nella finestra Modalità), si tolgono le modalità che non si ritengono coerenti con l'oggetto dell'analisi, ad esempio in questo caso sono state tolte le modalità "E' principalmente una crisi economica" e "Mi piace molto essere il capo". Si salvano le connessioni selezionate in un file di tipo testo (.txt) e a questo punto si può far leggere il file al fine di selezionare le variabili. L'obiettivo diventa quindi quello di studiare come le variabili correlate con "mi sento depresso" si strutturano in fattori.

Calcolare i fattori e ruotare gli assi fattoriali

Selezionate le variabili è possibile procedere al calcolo dei fattori. La tecnica utilizzata per il calcolo dei fattori è il metodo del centroide descritto da B. Cattell in "The scientific use of factor analysis". Effettuando l'analisi fattoriale sulle variabili che abbiamo prima selezionato si ottiene:



Si vede che le strutture individuate dalle frecce (verdi) sono a 90° tra loro e non possono perciò essere lette assieme (non sono correlate). Nell'esempio, le proiezioni di queste strutture sull'asse 1 (fattore 1) coincidono. Per evitare di leggere assieme strutture tra loro non correlate è quindi necessario procedere alla rotazione degli assi (come indicato nell'esempio).

Analisi dei risultati

I risultati delle analisi fattoriali vengono visualizzati riportando:

- la varianza spiegata da ogni fattore;
- la varianza delle variabili spiegata dai fattori;
- le correlazioni delle variabili con i fattori;

Per la lettura dei risultati, interessa in genere solo il terzo punto, cioè le correlazioni delle variabili dicotomiche con gli assi fattoriali. Vediamo comunque di descrivere tutti i risultati che vengono riportati dall'analisi fattoriale, facendo sempre riferimento all'esempio precedente:

----- **VARIANZA SPIEGATA** -----

Fat.	1:	0.506
Fat.	2:	0.178
Fat.	3:	0.174
Fat.	4:	0.142

Il totale della varianza spiegata dei fattori è uguale a 1.000. E' importante ricordare che la tecnica di analisi fattoriale qui utilizzata (analisi in fattori comuni, metodo del centroide) utilizza solo la parte comune delle variabili. La varianza spiegata è quindi relativa alla parte comune (che correla). Come si vede il primo fattore spiega oltre la metà della varianza, il secondo il 17,8%, il terzo il 17,4% e il quarto il 14,2%.

----- **VARIANZA DELLE VARIABILI** -----

45 -	0.830	F/2	Ho poca fiducia in me stesso
123 -	0.830	N/11	Mi sento molto impacciato
118 -	0.688	N/6	Vengo spesso criticato
119 -	0.681	N/7	Ho molta paura del giudizio degli altri
52 -	0.586	F/9	Mi sento brutto
47 -	0.574	F/4	Mi sento depresso
50 -	0.543	F/7	Ho poca fiducia nella vita
49 -	0.525	F/6	Mi sento solo
141 -	0.518	Q/5	Mi sento impacciato/a con la sessualità
53 -	0.512	F/10	Mi sento in grado di fare scelte per il mio futuro
46 -	0.449	F/3	Mi sento angosciato
51 -	0.429	F/8	Non mi piace il mio corpo
56 -	0.249	G/3	Che la vita non vale la pena di essere vissuta

Poiché solo la parte comune (la parte che correla) delle variabili viene tradotta in fattori, Sintropia-DS indica quanto di ogni variabile è spiegato dagli asse fattoriali.

Varianza generale: 0.570

La varianza generale indica quanta varianza delle variabili è spiegata dai fattori (fattori comuni). La parte restante viene chiamata "fattori singoli" ed indica la parte di varianza di

ciascuna variabile che non si è potuta spiegare tramite le correlazioni con le altre variabili.

Le correlazioni delle variabili con i fattori

A questo punto è possibile passare alla lettura ed interpretazione degli assi fattoriali. L'obiettivo è quello di dare un nome ad ogni fattore e di togliere dalla lista delle variabili quelle che possono essere entrate incidentalmente a causa di problemi legati allo schiacciamento dello spazio fattoriale su di un asse.

----- Fattore: 1 - **Fattore generale sulla depressione che unisce tutte le variabili**

1	0.757	(47)	F/4	Mi sento depresso
2	0.733	(50)	F/7	Ho poca fiducia nella vita
3	0.716	(49)	F/6	Mi sento solo
4	0.670	(46)	F/3	Mi sento angosciato
5	0.651	(51)	F/8	Non mi piace il mio corpo
6	0.490	(56)	G/3	Che la vita non vale la pena di essere vissuta
7	0.462	(45)	F/2	Ho poca fiducia in me stesso
8	0.422	(52)	F/9	Mi sento brutto
9	0.388	(118)	N/6	Vengo spesso criticato
10	0.388	(119)	N/7	Ho molta paura del giudizio degli altri
11	0.355	(53)	F/10	Mi sento in grado di fare scelte per il mio futuro
12	0.314	(141)	Q/5	Mi sento impacciato/a con la sessualità
13	0.309	(123)	N/11	Mi sento molto impacciato

----- Fattore: 2 - **Giudizio degli altri**

1	0.733	(118)	N/6	Vengo spesso criticato
2	0.728	(119)	N/7	Ho molta paura del giudizio degli altri
3	0.498	(123)	N/11	Mi sento molto impacciato

----- Fattore: 3 - **Sessualità e fisicità**

1	0.646	(141)	Q/5	Mi sento impacciato/a con la sessualità
2	0.639	(52)	F/9	Mi sento brutto
3	0.621	(53)	F/10	Mi sento in grado di fare scelte per il mio futuro
4	0.274	(123)	N/11	Mi sento molto impacciato

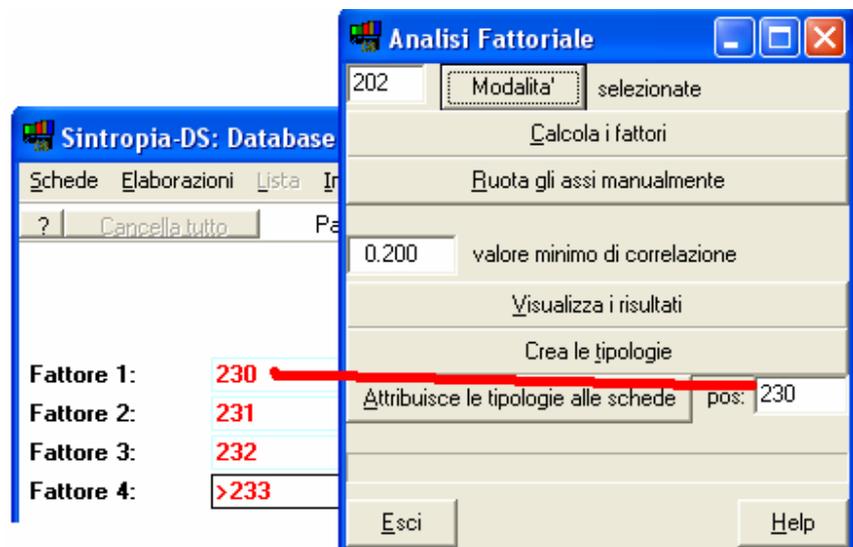
Come si vede, le variabili incoerenti con il contesto possono essere cancellate. E' importante sottolineare che l'analisi fattoriale è una guida, un aiuto alla lettura dei dati, ma non ha di per se un valore assoluto. In altre parole, solo quelle strutture alle quali si riesce ad attribuire un nome possono essere prese in considerazione. Strutture confuse o poco chiare possono dipendere da motivi incidentali legati all'analisi o alla mancanza di un numero sufficiente di variabili adatte a "chiarire" le strutture del fattore. Se non si riesce ad attribuire un nome all'asse fattoriale è necessario scartarlo. Assi che non si riescono a "leggere" non possono essere utilizzati ulteriormente.

Attribuzione dei punteggi fattoriali

Ad ogni scheda può essere assegnato un valore da 0 a 100 che indica la similarità con la struttura fattoriale. Ad esempio se il questionario presenta tutte le modalità della struttura fattoriale avrà valore 100 se non presenta modalità della struttura fattoriale avrà valore 0, ecc. In questo modo il fattore diventa una variabile, come tutte le altre, che può essere elaborato con le distribuzioni semplici, le tabelle doppie, le analisi correlazionali o inserito in nuove analisi fattoriali. Prima di effettuare l'attribuzione delle tipologie è necessario aver riservato sulla scheda tanti campi contigui quante sono le tipologie fattoriali che saranno attribuite. Le strutture fattoriali verranno lette direttamente dal file testo da voi salvato. Per ogni fattore da voi salvato verrà calcolata una tipologia.

Ad esempio, prendiamo il caso dei 3 fattori prima esaminati. Salviamo questi tre fattori in un file in formato .txt, depurando il testo da tutte le informazioni che non servono (ad esempio varianza spiegata, varianza delle variabili) e lasciando solo i fattori selezionati, con le variabili da noi selezionate. Sintropia-DS riconosce il passaggio da una tipologia all'altra in quanto ogni fattore inizia con dei trattini (----- fattore 1) è quindi importante lasciare questi riferimenti, altrimenti si rischia di far leggere tutte le variabili come appartenenti ad un'unica tipologia. Creato il file testo è necessario indicare la posizione della scheda dalla quale si inizia ad inserire i punteggi fattoriali. Per conoscere la posizione del campo utilizzate la finestra "Configura" (che trovate nel menù Schede).

Nella finestra relativa all'analisi fattoriale è necessario inserire il numero del primo campo riservato per i punteggi fattoriali. **Attenzione!!!** Il programma non ha modo di controllare se avete inserito un numero corretto o sbagliato. Quando premete il tasto "Attribuisce le tipologie" l'informazione presente su questi campi sarà sostituita con i punteggi fattoriali. Quindi, se avete inserito erroneamente la posizione di variabili diverse dai punteggi fattoriali perderete l'informazione che era presente in questi campi. E' perciò utile fare una copia dell'archivio prima di procedere nell'attribuzione delle tipologie.



Una volta attribuite le tipologie troverete i punteggi che variano da 0 a 100 a seconda della concordanza tra la scheda e la tipologia fattoriale. A questo punto è possibile trattare queste variabili come qualsiasi altra variabile della scheda. E' possibile inserire nella griglia le condizioni per la creazione di nuove variabili dicotomiche ed incrociare in questo modo i punteggi fattoriali con le variabili del questionario o vedere come questi si distribuiscono.

Ad esempio:

	n	%
Fattore 1 = 0%	55	5.65
Fattore 1 = 1-25%	225	23.10
Fattore 1 = 26-50%	255	26.18
Fattore 1 = 51-75%	224	23.00
Fattore 1 = 76-99%	193	19.82
Fattore 1 = 100%	22	2.26
Totale (schede)	974	



I punteggi fattoriali diventano delle nuove variabili a tutti gli effetti, variabili con le quali è possibile effettuare qualsiasi operazione, non ultimo nuove analisi fattoriali realizzate sui punteggi fattoriali.