

Elaborazioni

Ulisse Di Corpo

Il menù "Elaborazioni" consente di selezionare gruppi di schede, visualizzare conteggi e distribuzioni di frequenze, produrre liste/etichette, stampe personalizzate e consente di effettuare elaborazioni statistiche.

La prima opzione "Selezione delle schede" si utilizza per inserire condizioni di ricerca direttamente sulla scheda e per selezionare le schede che soddisfano le condizioni inserite. Sulle schede selezionate è poi possibile effettuare elaborazioni (analisi statistiche), liste, reports e rappresentazioni cartografiche.

E' possibile "Selezionare le schede" anche in altri due punti del programma:

1. nella sezione "Ricerca di schede" del menù "Schede";
2. nella sezione "Elaborazioni veloci" del menù Elaborazioni.

Quando si effettua una selezione di schede il risultato viene tenuto in memoria e viene salvato quando si esce dal programma o si cambia archivio. Le selezioni effettuate nelle Elaborazioni Veloci e nell'opzione "Selezione delle Schede" di questo menù possono interagire con selezioni effettuate precedentemente, aggiungendo condizioni di ricerca o sommandosi alle schede selezionate.

Sono disponibili due sezioni per le analisi statistiche dei dati:

1. *Elaborazioni: questa sezione utilizza criteri classici di lettura dell'archivio e i tempi di elaborazione sono proporzionali alle dimensioni dell'archivio.*
2. *Elaborazioni Veloci: questa sezione utilizza criteri di lettura dell'archivio che sono indipendenti dalle sue dimensioni. Ciò consente di effettuare elaborazioni istantanee indipendentemente dalle dimensioni dell'archivio.*

Selezione delle schede

Le condizioni di selezione delle schede vengono inserite direttamente sulla scheda. Se non sono state inserite condizioni di ricerca saranno selezionate tutte le schede, quando, invece, sono presenti condizioni di ricerca saranno selezionate solo le schede che soddisfano le condizioni inserite.

Che informazioni inserire sulla scheda:

1. posizionatevi sul campo nel quale volete inserire la condizione e premere il tasto "C", comparirà una finestra che consentirà di inserire la condizione. Se non dovesse comparire alcuna finestra significa che il campo è di tipo alfabetico, non codificato. In questa sezione, su campi alfabetici non è possibile inserire condizioni. Per inserire condizioni su campi alfabetici dovete utilizzare l'opzione Lista in "Ricerca di Schede" nel menù "Schede".
2. Per ricordarci della condizione inserita comparirà il carattere "=". Se volete cercare tutte le schede che non presentano la condizione inserita premete il tasto "#" o il tasto "=", comparirà il carattere "#".
3. Se volete calcolare dei conteggi (distribuzioni di frequenza), posizionatevi sul campo che desiderate interrogare e premere il tasto "S", comparirà l'indicazione "St", per ricordare che su quel campo saranno calcolati dei conteggi (statistiche).
4. Se volete calcolare delle distribuzioni statistiche quantitative potete selezionare la variabile da utilizzare come quantità e premere il tasto "Q". Per ricordarvi che tale variabile è stata selezionata come variabile quantitativa, comparirà il carattere Q.

Appena sotto la lista dei menù il programma presenta una barra con dei tasti e delle informazioni:



Come procedere nella selezione delle schede:

- premete il tasto "Ricerca". Questo tasto visualizza la finestra "Ricerca". Sono possibili 3 modalità di ricerca: *Prima ricerca*; *Aggiunge Ricerca*; *Legge Ricerca*. **Prima Ricerca** seleziona tutte le schede che soddisfano le condizioni di ricerca. **Aggiunge Ricerca** aggiunge alle schede selezionate le schede selezionate in una ricerca precedente, in questo modo è possibile effettuare un OR tra ricerche. **Legge Ricerca** seleziona, di una ricerca effettuata in precedenza, solo quelle schede che soddisfano le condizioni attualmente inserite, in questo modo si effettua un AND tra ricerche. N.B. se le elaborazioni veloci sono impostate, trovate in questa sezione delle opzioni di selezione di schede ottimizzate che risultano più facili da utilizzare e istantanee nei tempi di lettura dell'archivio.



Finita la ricerca potete visualizzare i conteggi posizionando il cursore sui campi contrassegnati con "St" e premendo il tasto "D" (distribuzione). Si tratta di un feed-back rapido che consente di verificare se la selezione effettuata è riuscita o ha incontrato dei problemi. I tasti "<<" e ">>" si utilizzano quando la distribuzione territoriale è stata calcolata su una variabile quantitativa

(ad esempio le percentuali) o continua (ad esempio le date). In questo modo è possibile stabilire l'intervallo da utilizzare per la visualizzazione della distribuzione delle frequenze.

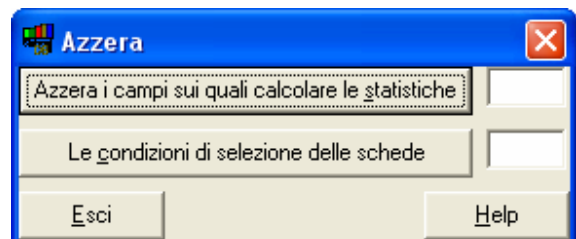
Questa sezione pone un limite nei conteggi. Quando le schede sono ampie e le modalità superano le 4.000 voci, i conteggi vengono calcolati per le prime 4.000 voci, le successive saranno omesse. Per sapere se una distribuzione è calcolabile, posizionatevi sulla variabile e premete il tasto "D". Se la variabile è collocata oltre il limite delle 4.000 modalità il programma avverte: "distribuzione non calcolabile".

Per accedere alle analisi dei dati propriamente dette (analisi statistiche) dovete selezionare le ultime opzioni del menù "Elaborazioni".

Effettuata una ricerca potete salvarla. A tal fine premete il tasto "Salva" dalla barra, comparirà la finestra "Salva" che consente di salvare la ricerca, le condizioni o la distribuzione territoriale. La distribuzione territoriale viene prodotta quando si fanno calcolare i conteggi (St) su un campo che utilizza la codifica ufficiale dei comuni ISTAT. La distribuzione territoriale, una volta salvata, può poi essere letta da Sintropia-AS ed essere visualizzata ed elaborata come variabile territoriale.



Quando effettuate una nuova ricerca e se state utilizzando schede di più pagine è bene azzerare le condizioni di selezione delle schede. Può infatti capitare che sfuggano delle condizioni di ricerca e che le condizioni inserite si vadano a sommare a condizioni già presenti. E' quindi buona prassi azzerare le condizioni di ricerca prima di inserirne di nuove. A tal fine si utilizza il secondo tasto "Azzerà".



Allo stesso modo del salvataggio è possibile recuperare ricerche e condizioni di selezione delle schede ed è possibile caricare selezioni territoriali memorizzate Sintropia-AS. Quest'ultima opzione, abbinata con l'opzione precedente di salvataggio della distribuzione territoriale è particolarmente importante. Infatti, da una parte è possibile salvare i risultati sottoforma di distribuzioni territoriali visualizzabile per mezzo di cartine ed analisi territoriali, dall'altra è possibile recuperare le selezioni territoriali effettuate da Sintropia-AS come risultato delle analisi territoriali.



Lista le schede selezionate

Questa sezione consente di effettuare liste sequenziali, ordinate alfabetiche e ordinate in base all'interazione di più livelli.

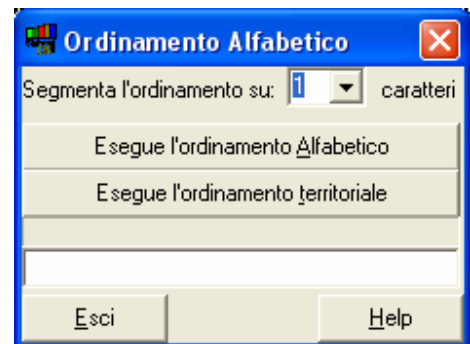
Appena sotto la linea dei menù il programma presenta una barra con dei tasti e delle informazioni:



- il primo tasto “?” consente di accedere all’help. *In tutti i menù del programma è presente un tasto help o “?”.*
 - Il secondo tasto “Maschera” consente di stabilire quali variabili saranno stampate. *Come già detto, le schede possono essere molto ampie e l’interesse può quindi essere quello di stampare solo parte del contenuto. Questa finestra consente di “Creare una maschera di stampa” nella quale sono riportate tutte le variabili della scheda, “Visualizzare” e modificare il contenuto della maschera di stampa, “Salvare” le maschere e “Aprire” maschere precedentemente salvate.*
-
- Il terzo tasto “Liste” consente di effettuare le “Liste” che saranno memorizzate nel file *liste.txt* (nell’indirizzo del programma). *Le liste possono essere effettuate utilizzando l’ordinamento alfabetico e l’ordinamento sequenziale e suddividendo la lista in gruppi in base alla codifica riportata su una o più variabili codificate (livelli). Per effettuare liste ordinate alfabeticamente dovete indicare sulla scheda il campo da utilizzare per l’ordinamento (per attivare il campo utilizzate il tasto “o”) ed effettuare l’ordinamento utilizzando il tasto “Ordinamenti”; solo dopo aver effettuato queste operazioni potete produrre la lista ordinata alfabeticamente. La lista può anche essere suddivisa in livelli, ad esempio divisa per provincia e all’interno per sesso. I campi da utilizzare per effettuare la suddivisione in livelli sono selezionati con il tasto “I”. In tutto è possibile far interagire due campi più un campo per l’ordinamento alfabetico.*
-
- Il terzo tasto “Vede Lista.Txt” consente di visualizzare la lista che è stata prodotta. E’ importante far presente che cliccando su questa lista il programma si sposta automaticamente sulla scheda corrispondente alla riga sulla quale si è cliccato il mouse. In questo modo è possibile creare una interazione tra lista e schede del programma, potendo saltare in continuazione dalla lista alle schede (per tornare alla lista basta premere il menù “Lista”).
 - Il quarto tasto “Ordinamenti” visualizza la finestra “Ordinamento Alfabetico”, in questa finestra è possibile **Eseguire l’ordinamento alfabetico**: *Effettua l’ordinamento alfabetico del campo selezionato con il tasto O. Poiché gli archivi possono essere molto estesi, l’ordinamento è suddiviso in pacchetti in base al primo carattere o a più caratteri iniziali del campo. E’ necessario indicare su quanti caratteri del campo suddividere i pacchetti da ordinare alfabeticamente.*

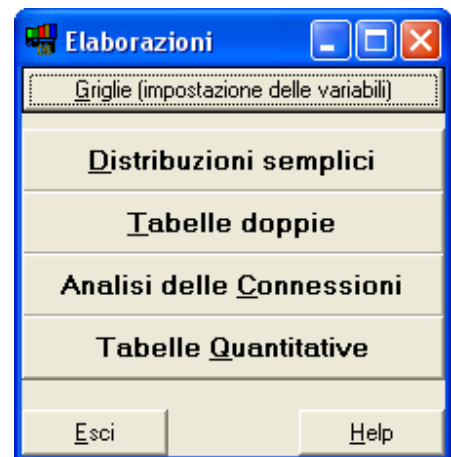
Ad esempio, normalmente l'ordinamento si effettua suddividendo l'archivio in base al primo carattere, ciò significa che tutte le schede che afferiscono alla lettera "A" saranno trattate come un pacchetto, e così via per la lettera "B", "C" ecc., ma prendiamo il CAP, per effettuare l'ordinamento alfabetico dei CAP, ad esempio a ROMA sono presenti molte schede, i primi tre caratteri del CAP non saranno utili per segmentare l'archivio in pacchetti, è quindi necessario suddividere l'archivio prendendo i primi 4 caratteri dei CAP. La suddivisione in pacchetti nasce dal fatto che il programma non può ordinare pacchetti più ampi di 10.000 schede. L'ordinamento alfabetico riconosce la differenza tra lettere maiuscole e minuscole.

- Il tasto "Cancella", cancella la selezione del campo utilizzato per l'ordinamento alfabetico e dei campi utilizzati per dividere la lista in gruppi.



Elaborazioni

Sintropia-DS elabora i dati su una scheda "virtuale" in cui tutte le informazioni sono tradotte in Si/No (variabili dicotomiche). Qualsiasi informazione nominale, ordinale, a intervalli o a rapporti può essere tradotta in una o più variabili dicotomiche. La **griglia** indica come realizzare questa traduzione. Prima di poter accedere alle elaborazioni dei dati è necessario impostare la Griglia. Le variabili dicotomiche (*si/no, vero/falso, presente/assente*) sono il livello più semplice di misura. Tutte le informazioni, anche quelle quantitative, possono essere tradotte in una o più variabili dicotomiche. Con le variabili dicotomiche diventa possibile far interagire qualsiasi tipo di informazione: informazioni quantitative e qualitative, variabili nominali, ordinali, a intervalli e a rapporti.



Analisi basate su variabili dicotomiche consentono analisi globali che permettono di ricostruire la complessità dei fenomeni. Un po' come capita con l'informatica, dal bit, dal si/no, dal livello più semplice, è possibile ricostruire qualsiasi tipo di complessità. E' diffuso il pregiudizio che il trattamento dell'informazione in base a variabili dicotomiche porti a perdere informazione, in quanto le variabili dicotomiche sono il livello più basso di misura e sembrerebbe quindi un controsenso tradurre le informazioni quantitative in variabili dicotomiche. Anche in campo informatico si cerca di costruire computer analogici che al posto di informazioni dicotomiche (si/no) utilizzano informazioni quantitative. Ma, finora, questi tentativi sono falliti. Lo stesso accade per l'analisi dei dati: perché si possa arrivare ad analisi in grado di valorizzare la complessità dei fenomeni è necessario partire dal livello più semplice di informazione, le variabili dicotomiche o come alcuni le chiamano i bit di informazione.

Come si traducono le informazioni in variabili dicotomiche.

Le informazioni vengono in genere raccolte sottoforma di variabili:

1. nominali
2. ordinali
3. ad intervalli
4. a rapporti

Queste 4 categorie di variabili si distinguono per le operazioni matematiche, sempre più complesse, che possono essere utilizzate:

1. nel caso di variabili *nominali* il valore è una modalità, ad esempio la provincia, il colore degli occhi, lo stato civile. Con le variabili nominali è possibile effettuare solamente conteggi (contare il numero di risposte per ogni singola modalità);
2. le variabili *ordinali* si differenziano dalle variabili nominali in quanto, oltre a consentire l'analisi dei conteggi (delle frequenze), consentono di effettuare analisi basate sull'ordine delle modalità. Ad esempio, con la variabile ordinale *istruzione* oltre a contare quante sono le persone per ogni modalità, (elementare, media, superiore e universitaria), è possibile contare qual è il livello di istruzione raggiunto da almeno (ad esempio) il 50% della popolazione, ecc... In altre parole esiste un ordine, una progressione. Si può utilizzare quest'ordine per creare nuove categorie (ad esempio: istruzione bassa e istruzione elevata) o per ordinare la popolazione (prima i laureati, poi i diplomati, quindi i possessori di licenza media, ecc...);
3. le variabili ad *intervalli* si differenziano dalle variabili ordinali in quanto, oltre a consentire l'analisi dei conteggi e l'ordinamento della popolazione, introducono il concetto di misura: tra ogni modalità contigua esiste la stessa distanza, la stessa misura. Ad esempio, la distanza tra 1 grado e 2 gradi è la stessa che si trova tra 2 gradi e 3 gradi. Le osservazioni a intervalli possono essere analizzate oltre che con conteggi e ordinamenti, anche con operazioni quali l'addizione, la sottrazione ed è possibile calcolare i valori medi;
4. le variabili a *rapporti* si differenziano da quelle a intervalli per la coincidenza dello 0 con lo 0 assoluto. Ad esempio non è possibile dire che 2 gradi centigradi siano il doppio di 1 grado centigrado. Infatti 0 gradi centigradi non corrispondono allo zero assoluto, ma alla temperatura in cui l'acqua si trasforma in ghiaccio. Lo zero assoluto corrisponde a -273 gradi centigradi (0 gradi Celsius). Quindi 1 grado centigrado nei fatti corrisponde a 274 gradi Celsius, 2 gradi corrisponde a 275 gradi Celsius: 275 gradi Celsius non è il doppio di 274 gradi.

La traduzione di queste variabili in variabili dicotomiche segue alcuni accorgimenti:

- nel caso di variabili nominali, la singola modalità (ad esempio la singola provincia, il singolo colore) può diventare una variabile dicotomica. Ad esempio la regione *Lazio* diventa la variabile dicotomica *Lazio* per la quale le risposte possono assumere solo i valori si/no. Durante la fase di progettazione della scheda con la quale si rilevano i dati è utile evitare di essere troppo analitici e puntuali. Ad esempio, se si effettua una ricerca raccogliendo 200 questionari e questa ricerca è distribuita su tutto il territorio nazionale, non ha senso chiedere il comune di residenza. Infatti, essendo i comuni italiani più di 8.000, la frequenza di risposta per i singoli comuni sarà eccessivamente bassa. E' invece necessario individuare che cosa interessa o è possibile sapere della variabile territorio, ad esempio Nord/Centro/Sud o dimensione del comune, ecc... ed inserire nel questionario direttamente queste domande. Risulta utile raccogliere l'informazione nello stesso modo in cui servirà per le elaborazioni. Le ricodifiche dei dati, la loro trasformazione, sono processi che possono risultare troppo onerosi e che possono portare al mancato utilizzo di variabili raccolte in forma troppo dettagliata.
- Le variabili ordinali sono tutte quelle in cui le modalità di risposta seguono una progressione, ad esempio il titolo di studio o variabili del tipo mai, poco, abbastanza, molto. Queste variabili possono essere trattate allo stesso modo delle variabili nominali traducendo ogni singola modalità di risposta in una variabile dicotomica. Ma, è anche possibile tradurre l'informazione nella forma alto/basso, poco/molto, ad esempio titolo di studio alto/basso. E' importante sottolineare che non esiste un criterio oggettivo per delimitare l'alto dal basso. In ogni studio le modalità che separeranno l'alto dal basso possono variare. Ad esempio se si effettua uno studio sui professori universitari il titolo di studio basso potrà essere la laurea, mentre il titolo di studio alto le specializzazioni. Se si effettua uno studio tra la popolazione povera di un paese del terzo mondo, la separazione tra alto e basso corrisponderà a scolarizzato e non scolarizzato. Questo esempio mostra che la divisione di una variabile ordinale in una variabile dicotomica alto/basso, deve sempre tener conto del contesto e delle finalità dello studio. Nel caso in cui le informazioni che si chiedono sono tali da non consentire una suddivisione ragionata tra alto e basso, si suggerisce di equiripartire la popolazione tra le modalità alto/basso. Ad esempio, se si deve dicotomizzare la scala *mai, poco, abbastanza, molto*, il punto di separazione tra l'alto e il basso sarà quel punto che consente il massimizzare il bilanciamento dell'alto con il basso (valore mediano).
- Per le variabili continue (variabili ad intervalli e variabili a rapporti) si utilizzano in genere dei valori di taglio (cut-off points) che delimitano il passaggio da valori bassi a valori alti (da valori normali a valori patologici) o delle classi, degli intervalli. E' importante sottolineare che per queste variabili il taglio tra valori bassi e valori alti non segue necessariamente il criterio dell'equiripartizione. Infatti, queste variabili hanno un valore quantitativo preciso ed in genere a queste variabili si associano valori critici che indicano il passaggio da uno stato all'altro (ad esempio il passaggio dalla salute alla patologia). L'obiettivo del ricercatore e lo scopo dell'analisi dei dati è in genere quello di individuare questi valori critici. Capita quindi frequentemente che una stessa variabile continua venga tradotta in più variabili dicotomiche al fine di individuare quale valore di taglio consente meglio di altri di individuare il passaggio da uno stato all'altro.

Griglia di traduzione delle variabili in variabili dicotomiche

Sintropia-DS produce automaticamente la griglia partendo dal file *scheda.txt* in cui è stato memorizzato il tracciato record. E' comunque possibile intervenire in qualsiasi momento sulla griglia, ad esempio per modificare delle classi (come le classi di età, ecc.). Ogni riga della griglia definisce una variabile dicotomica. La variabile dicotomica è composta da una sigla, dal riferimento alla variabile originaria, dall'intervallo che la determina e da alcune informazioni che indicano se è la prima variabile di un gruppo, se fa parte di una domanda a risposta multipla, se si tratta della prima variabile di una tabella sinottica, se le percentuali debbano essere calcolate su di una particolare variabile filtro.



Ad esempio:

- 1. Attivo*1=1,1;G0S0
- 1. Cancellato*=2,2;
- 1. MR*=0,0;

La prima variabile dicotomica **1. Attivo** è prodotta dalla variabile ***1**, ed è vera nel caso in cui il valore cada nell'intervallo **1,1**. E' la prima variabile di un gruppo **;G** non fa parte di una variabile a risposta multipla (**0**, altrimenti si sarebbe inserito il numero di risposte, non fa parte di una variabile sinottica (**S**, altrimenti si sarebbe utilizzato il carattere **s**), non fa riferimento ad una variabile filtro (**0**).

La seconda variabile dicotomica **2. Cancellato** è stata realizzata considerando la stessa variabile (non si ripete quindi il numero della variabile), l'intervallo che la rende vera è **2,2**, non è la prima variabile di un gruppo.

La terza variabile dicotomica **1. MR** indica la mancata risposta. Quando si trattano variabili codificate (che fanno riferimento a liste) lo zero è la mancata risposta, nel caso di variabili quantitative (ma anche di ore e date) la risposta nulla è -32.000. Nel caso di variabili quantitative lunghe è -9.999.999.

Per creare la griglia utilizzate il tasto "Crea la griglia". Il programma creerà i file *grl-sch.txt* e *tin-sch.txt*. Il primo con la griglia, il secondo con le intestazioni delle tabelle. La griglia non può superare le 4.000 righe. Sintropia-DS avverte quando si supera questa soglia. Nel caso si dovesse lavorare con un numero di modalità superiore alle 4.000 righe è necessario scomporre la griglia in due o più griglie.

Nel momento in cui modificate la griglia può capitare di cancellare intenzionalmente delle intere tabelle o inserirne delle nuove (ad esempio non volete la lista dettagliata delle province

ma volete la suddivisione Nord, Centro, Sud). A questo punto è importante modificare anche il file *tin-sch.txt* nel quale sono riportate le intestazioni delle tabelle. L'intestazione di una tabella è inserita nel modo seguente:

XXX <- numero progressivo della tabella

xxxxxxxxxxxxxxxxxxxxxxx <- intestazione (anche su più righe)

/// <- chiusura

Ad esempio:

```
001

1. Operatore:
///
002

2. Data:
///
003

3. Ora Inizio:
///
004

5. Provincia:
///
005

6. Nazione:
///
```

Ovviamente togliendo o aggiungendo una tabella è necessario rinumerare tutte le intestazioni. A tal fine trovate l'opzione "Rinumera Intestazioni".

Quando il file *grl-sch.txt* supera i 32K (32.000 caratteri) è necessario effettuare le modifiche utilizzando una video-scrittura esterna in modalità testo ".txt". Le finestre di Windows utilizzate dal programma non consentono, infatti, di lavorare con file superiori a 32K. Per evitare problemi di conflitto di condivisione dello stesso file consigliamo di utilizzare "WordPad" e di aprire e salvare il file *grl-sch.txt* come file di testo (.Txt).

Modificando il file *grl-sch.txt* capita spesso di inserire degli errori, ad esempio manca il punto e virgola, non è definito bene l'intervallo in cui una variabile dicotomica è vera, ecc... Il tasto

“Diagnostica” consente di verificare la presenza di errori. E' bene utilizzare questa opzione prima di avviare le elaborazioni dei dati, le procedure di elaborazione dati non diagnosticano infatti gli errori nella griglia e potrebbero perciò portare a degli errori di esecuzione.

Griglia per elaborazioni quantitative

Con le variabili quantitative è possibile calcolare la somma, la media, la varianza, valori minimi e massimi.

Con la griglia quantitativa si indicano le variabili da trattare come variabili quantitative, la loro sigla, e la loro eventuale appartenenza a gruppi.

Ad esempio:

```
33*medici di ruolo*SG
34*infermieri di ruolo*s
35*assistenti sociali di ruolo*s
36*medici volontari*SG
37*infermieri di volontari*s
38*assistenti sociali volontari*s
```

Ogni riga corrisponde ad una variabile quantitativa. La prima informazione riporta il numero della variabile, la seconda la sigla, la terza lo stato di tabella sinottica e di gruppo.

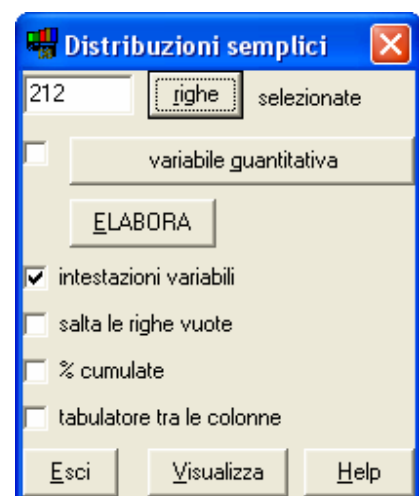
Ad esempio il carattere S indica l'inizio di una tabella sinottica, mentre s indica la continuazione e G la suddivisione in gruppi (righe).

Distribuzioni semplici

Per calcolare le distribuzioni semplici basta selezionare le righe, elaborare e visualizzare. E' possibile effettuare distribuzioni pesando ogni scheda in base ad una variabile quantitativa. In questo caso si otterranno distribuzioni di quantità.

Le opzioni presenti in questa sezione sono:

- intestazioni variabili. *Le variabili dicotomiche sono suddivise in gruppi (tabelle), per ogni gruppo è possibile specificare una intestazione (file tin-sch.txt, vedi griglia).*
- salta le righe vuote. *Per rendere le tabelle più sintetiche è possibile far saltare le righe (le modalità) che non ottengono risposte.*
- % cumulate. *Le percentuali possono essere calcolate come rapporto delle risposte ottenute da ogni singola variabile dicotomica sul totale o cumulando (sommando) le*



percentuali di ogni variabile a quelle precedenti (percentuali cumulate).

- tabulatore tra le colonne. Per facilitare l'acquisizione da parte di altri programmi delle tabelle prodotte dal programma può a volte risultare utile la separazione delle colonne per mezzo del carattere di tabulazione.

Selezione delle righe

Questa opzione consente di selezionare le variabili su cui effettuare le elaborazioni. Variabili che possono essere selezionate:

- attivandole direttamente con il mouse;
- attivando/disattivando la prima e trascinando il cursore premendo il tasto "Shift" (maiuscolo) e le frecce verticali della tastiera;
- selezionando il titolo della tabella e attivando/disattivando tutte quelle variabili che si riferiscono a tale tabella;
- utilizzando i tasti Attiva e Disattiva.

In genere si attivano tutte le variabili dicotomiche, si elabora e si visualizzano i risultati. Altre volte interessano solo

alcune variabili e l'attivazione di queste può in certi casi richiedere un certo lavoro. Per questo motivo sono stati inseriti i tasti Salva e Recupera che consentono di Salvare e Recuperare le configurazioni di attivazione delle variabili dicotomiche.

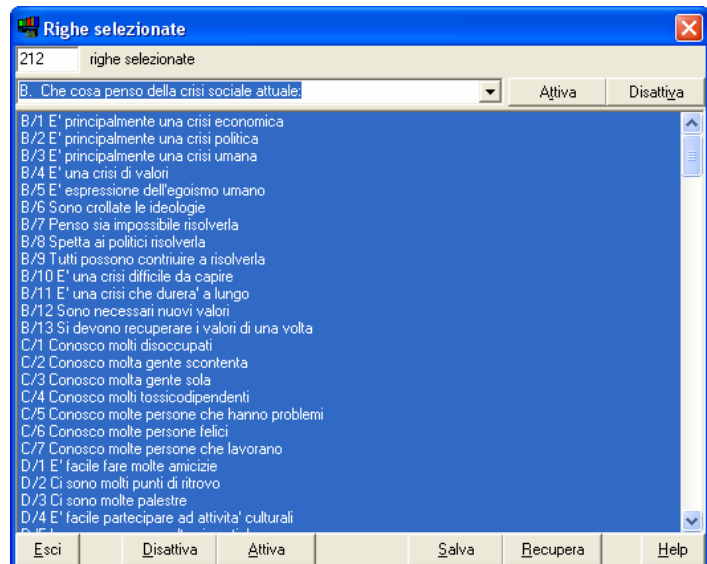


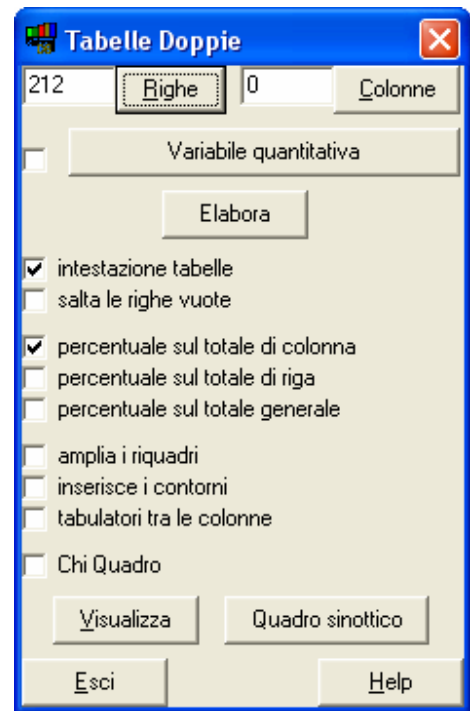
Tabelle doppie

Per calcolare le tabelle doppie è necessario selezionare le righe e le colonne da incrociare.

E' possibile effettuare le tabelle pesando ogni scheda in base ad una variabile quantitativa. In questo caso si otterranno tabelle di quantità. Le opzioni presenti in questa sezione sono:

- intestazione tabelle. Le variabili sono suddivise in gruppi (tabelle), per ogni gruppo è possibile specificare una intestazione (file tin-sch.txt, vedi griglia).
- salta le righe vuote. Per rendere le tabelle più sintetiche è possibile saltare le righe che non ottengono risposte.
- percentuale sul totale di colonna. L'obiettivo primario delle tabelle doppie è quello di confrontare le colonne che sono state selezionate. Il confronto viene effettuato utilizzando le percentuali di colonna.
- percentuale sul totale di riga. In alcuni casi l'interesse è quello di confrontare le righe. Consigliamo di cercare, quando possibile, di utilizzare la lettura standard che è quella del confronto tra percentuali di colonna.

- percentuale sul totale generale. *In alcuni casi l'interesse è quello di confrontare le percentuali che si ottengono utilizzando quale riferimento il totale generale della tabella.*
- amplia i riquadri. *Quando si realizzano tabelle di quantità, può essere utile ampliare la dimensione dei riquadri.*
- inserisce i contorni. *Vengono inseriti, in formato testo, dei contorni alle celle della tabella.*
- tabulatore tra le colonne. *Per facilitare l'acquisizione dei dati da parte di alcuni software è possibile introdurre, tra le colonne, un carattere di tabulazione.*
- Chi Quadrato. *Per valutare se le differenze che si osservano nelle tabelle sono statisticamente significative è possibile visualizzare l'indice del Chi Quadrato. Quando il valore del Chi Quadrato è seguito da un asterisco (*) la significatività statistica è del 5%, quando è seguito da due asterischi (**) la significatività statistica è dell'1%, quando compaiono 3 asterischi la significatività statistica è dell'1/1000. Se non compaiono asterischi, la tabella non presenta differenze statisticamente significative. E' importante sottolineare che il Chi Quadro viene calcolato sul complesso della tabella, spesso l'incrocio di singole variabili dicotomiche risulta significativo, ma quando la significatività si calcola sul complesso della tabella questa significatività si perde nell'insieme degli incroci (ogni singola cella di una tabella è un incrocio di due variabili dicotomiche) perdendo di significatività.*



Ad esempio:

col. 1 = Maschio
col. 2 = Femmina

B. Che cosa penso della crisi sociale attuale:

	1	2	Totale
B/1 E' principalmente una crisi	24	44	76
% Col.	5.84	8.53	7.80
B/2 E' principalmente una crisi	193	226	439
% Col.	46.96	43.80	45.07
Totale (schede)	411	516	974

Le prime righe:

col. 1 = **Maschio**
col. 2 = **Femmina**

riportano le sigle delle variabili dicotomiche in colonna, in questo esempio nella prima colonna abbiamo i maschi, nella seconda le femmine.

Segue l'intestazione della tabella (che potete modificare intervenendo sul file *tin-sch.txt*):

B. Che cosa penso della crisi sociale attuale:

Segue l'intestazione delle colonne. Sintropia-DS contrassegna le colonne con il numero di colonna, ad esempio:

1 2 Totale

per sapere a che cosa questi numeri corrispondono andate all'inizio del testo dove trovate le sigle delle variabili di colonna. Seguono gli incroci:

<i>B/1 E' principalmente una crisi</i>	24	44	76
% Col.	5.84	8.53	7.80
<i>B/2 E' principalmente una crisi</i>	193	226	439
% Col.	46.96	43.80	45.07

le sigle delle variabili in riga vengono troncate per dar spazio alla tabella. Può quindi risultare utile intervenire su queste sigle, modificando il file *grl-sch.txt*, per renderle più leggibili. Nelle connessioni non si presenta questo problema di sigle troncate (in quanto le sigle seguono la tabella). Spesso nasce l'esigenza di utilizzare due file *grl-sch.txt*, un file che riporta le sigle sintetizzate/troncate da utilizzare con le tabelle e le distribuzioni, un file che riporta le sigle per esteso adatto per la lettura delle connessioni. Potete ovviamente organizzare più versioni del file *grl-sch.txt* (utilizzando i tasti salva e recupera che trovate nella finestra griglia).

La tabella viene, infine, chiusa dalla riga del totale:

<i>Totale (schede)</i>	411	516	974
------------------------	-----	-----	-----

Tra parentesi (**schede**) ricorda che il totale è relativo al numero di schede conteggiate, non al totale eseguito sommando le righe della tabella. Questa differenza è importante in quanto nei casi in cui la somma delle righe della tabella diverge dal totale delle schede è legittimo calcolare le percentuali solo sul totale delle schede.

Analisi delle connessioni

Le connessioni consentono di incrociare ogni variabile dicotomica con le altre variabili dicotomiche della scheda. In questo modo l'analisi delle connessioni:

- **evidenzia la tipologia** di ogni variabile dicotomica;
- **individua le differenze** significative e quindi le tabelle doppie informative. In altre parole, le connessioni danno in modo diverso, la stessa informazione che si ottiene dalle tabelle doppie. Le connessioni sono calcolate per tutti i singoli incroci, le singole caselle delle tabelle e se si presentano differenze significative (nelle percentuali di colonna delle tabelle) parallelamente si presentano connessioni.

Per calcolare le connessioni è necessario selezionare le variabili colonna (o variabili indipendenti) e le variabili riga (o variabili dipendenti). Le connessioni sono un modo sintetico di lettura delle tabelle. In altre parole, le selezioni effettuate per il calcolo delle tabelle vengono riproposte in questa sezione (e viceversa). Se avete già selezionate delle variabili in colonna e delle variabili in riga per le tabelle, potete mantenere questa selezione senza effettuare modifiche. Selezionate le righe e le colonne potete quindi Elaborare. *Se avete effettuato l'Elaborazione per le tabelle (con le stesse righe e colonne) non è necessario rifeffettuarla in questa sezione.*



Per la visualizzazione potete scegliere se visualizzare le relazioni banali e qual è il valore minimo di Chi Quadrato che volete considerare.

Il valore 6,635 del Chi Quadrato corrisponde ad una significatività statistica dell'1% e viene proposto automaticamente da Sintropia-DS. Il valore 3,841 corrisponde ad una significatività statistica del 5%, mentre il valore 10,827 corrisponde ad una significatività statistica dell'1‰ (un per mille). Se avete bisogno di cercare altri valori potete consultare in un qualsiasi manuale di statistica la distribuzione del Chi Quadrato, tenendo conto che quando si incrociano variabili dicotomiche i gradi di libertà corrispondono ad 1.

Le relazioni banali sono quelle relazioni che si creano automaticamente tra variabili dicotomiche che appartengono alla stessa tabella. Ad esempio maschi è correlato negativamente a femmine, licenza elementare è correlato negativamente con licenza media, diploma superiore e laurea... in altre parole l'appartenenza ad una variabile dicotomica porta automaticamente alla non appartenenza all'altra variabile dicotomica.

A questo punto potete visualizzare i risultati.

Come leggere i risultati delle connessioni

ChiQ	rPhi	(N)	(% / %)	Discordanze
974.00	1.000	(147)	(100.00% / 0.00%)	Maschio
272.92	0.529	(138)	(92.94% / 41.21%)	Q/2 Mi sento molto timido/a in coppia
42.22	0.208	(114)	(69.34% / 48.49%)	N/2 Sono molto affezionato ai miei amici
30.50	0.177	(143)	(46.96% / 29.66%)	Q/7 Con mia madre posso confidarmi in merito alla sessualità
28.53	0.171	(142)	(99.76% / 92.72%)	Q/6 Cambio partner molto spesso
23.04	0.154	(144)	(99.76% / 93.96%)	Q/8 Con mio padre posso confidarmi in merito alla sessualità
14.41	0.122	(15)	(42.34% / 30.55%)	C/2 Conosco molta gente scontenta

Annotations in the image:
 - 'Variabile' points to the 'Maschio' header.
 - 'Concomitanze' points to the 'Maschio' header.
 - 'Discordanze' points to the 'Maschio' header.
 - 'Indice del Chi Quadrato' points to the 'ChiQ' column.
 - 'Indice r.Phi' points to the 'rPhi' column.
 - 'Numero d'ordine che identifica la variabile' points to the variable names in the last column.
 - 'Variabili righe' points to the last column.

La prima riga riporta la sigla della variabile selezionata in colonna, in questo esempio la variabile "Maschio", le righe successive riportano le variabili selezionate in righe, ordinate in base alla significatività statistica dell'incrocio. La prima riga riporta la variabile che presenta il massimo valore di connessione con la variabile "Maschio". Come si può vedere nell'esempio si tratta di una relazione "banale" in quanto incrociando la variabile "Maschio" con se stessa il valore di connessione è massimo ($r\Phi = 1.000$, Chi Quadrato = numero di schede conteggiate), l'incrocio della variabile colonna con la variabili riga ottiene una concordanza pari al 100% (cioè nel 100% dei casi quando si presenta la modalità maschio come variabile dipendente si presenta anche la modalità maschio come variabile indipendente) e una discordanza pari allo 0%.

La seconda riga presenta invece l'incrocio tra la variabile indipendente "Maschio" e la variabile dipendente "Mi sento molto timido in coppia":

272.92 0.529 (138) (92.94%/ 41.21%) Q/2 Mi sento molto timido/a in coppia

Il 92.94 dei maschi afferma "Mi sento molto timido in coppia" mentre il 41,21% dei non maschi afferma "Mi sento molto timido in coppia". Questa diversità tra le due percentuali porta ad un Chi Quadrato di 272,92 e ad un $r\Phi$ di 0.529. La relazione è statisticamente significativa in quanto il Chi Quadrato supera il valore 6.635 (significatività dell'1%). Si può quindi affermare che i maschi sono caratterizzati da una maggiore timidezza in coppia.

Connessioni e tabelle doppie

Ovviamente, i valori in percentuale che le tabelle delle connessioni presentano devono poter essere ritrovati nelle tabelle doppie. Le connessioni sono infatti un modo diverso e più sintetico di leggere la stessa informazione che viene prodotta dalle tabelle doppie.

Ad esempio la tabella in cui si incrocia la variabile sesso con la variabile "Mi sento molto timido/a in coppia" mostra:

Q. Nel rapporto di coppia:

	Maschi	Femmine	MR	Totale
Q/2 Mi sento molto timido/a in	382	199	33	614
% Col.	92.94	38.57	70.21	63.04
Totale (schede)	411	516	47	974

Come si vede sommando le femmine con la mancata risposta ($199 + 33 = 232$) e dividendo questo valore per il loro totale ($516 + 47 = 563$) porta al valore 0,41207 pari ad una percentuale di 41,21%. Otteniamo quindi dalle tabelle doppie gli stessi valori in percentuale osservati nelle connessioni: 92,94% per i maschi contrapposto a 41,21% dei non maschi.

Numerosità minima delle variabili

Si possono presentare valori del Chi Quadrato e del rPhi elevati anche quando la frequenza di una variabile è bassa. Ad esempio, mettiamo il caso che nell'esempio precedente invece di 411 maschi avessimo un solo maschio. Incrociando la variabile "Maschio" con se stessa si sarebbe egualmente osservato un valore del Chi Quadrato pari a 974 (numero delle schede) e un valore del rPhi pari a 1.000. Ciò evidenzia il fatto che valori di connessione elevati possono avere un significato ben diverso. Quando i valori di connessione si ottengono con il contributo di pochi soggetti è bene "prenderli con le pinze", quando i valori di connessione si ottengono con il contributo di molti soggetti il significato è sicuramente forte.

Numerosità minima del campione

Affinché i valori di connessioni possano essere presi in considerazione, senza incorrere in errori dovuti a eventi singoli o casi singoli, è importante che le variabili effettivamente varino. A tal fine è convenzione scartare tutte quelle variabili dicotomiche che presentano frequenze inferiori ai 30 casi. Ciò significa che per effettuare una analisi delle connessioni il campione utilizzato deve essere comunque superiore ai 60 casi (condizione che in campo medico, quando si lavora su patologie rare, può risultare problematico).

Numerosità del campione e precisione dello strumento

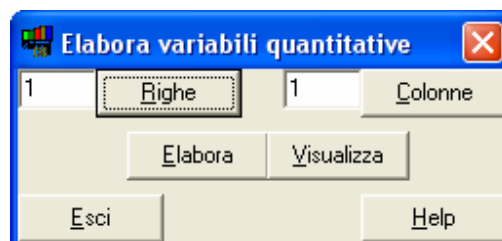
Maggiore è la numerosità dei casi, maggiore è la precisione delle relazioni. Ad esempio quando si utilizzano 100 soggetti e si cercano valori significativi all'1% (Chi Quadrato pari a 6,635), il valore di significatività del rPhi inizierà da $\sqrt{6,635/100}$, cioè 0,257, mentre quando si utilizzano 200 soggetti il valore di significatività dell'rPhi inizierà da $\sqrt{6,635/200}$ cioè 0,182. In altre parole, più è numeroso il campione, maggiore è la sensibilità delle connessioni. Ciò è importante in quanto spesso le relazioni più interessanti non sono quelle più forti, che in genere sono abbastanza ovvie, ma sono quelle intermedie e basse che richiedono perciò una numerosità del campione più elevata.

Caratteristiche del campione

Come visto precedentemente, al fine di realizzare studi basati sulle connessioni il pre-requisito è quello di massimizzare la variabilità delle variabili. In altre parole invece di campioni rappresentativi (randomizzati), i campioni devono essere "eterogenei". Ad esempio, se l'obiettivo dell'indagine è quello di studiare cosa si lega ad una determinata patologia sarà necessario organizzare il campione in modo da avere il 50% dei casi con la patologia e il 50% senza la patologia. L'obiettivo delle analisi delle connessioni è quello di trovare connessioni e di individuare successivamente gli ambiti di validità di tali connessioni. E' errato pensare che l'obiettivo sia, invece, quello di generalizzare l'informazione a tutta la popolazione.

Tabelle quantitative

Per ogni variabile quantitativa selezionata (in Riga) viene calcolata la media, il valore minimo, il valore massimo e lo scarto. I dati possono essere scomposti in base a variabili dicotomiche selezionate come variabili di colonna (ad esempio Nord, Centro, Sud).



Ad esempio, si prenda il caso dei punteggi fattoriali prodotti nel punto precedente, selezionando questi punteggi fattoriali come righe e il sesso come colonne si ottiene:

Fattore 1:	Totale	Maschi	Femmine
Totale.....:	45258	17297	25668
Rispondenti.....:	974	411	516
Media sui rispondenti:	46.47	42.09	49.74
Media sul totale.....:	46.47	42.09	49.74
Scarto.....:	29.01	28.85	28.76

Fattore 2:	Totale	Maschi	Femmine
Totale.....:	44634	17066	24977
Rispondenti.....:	974	411	516
Media sui rispondenti:	45.83	41.52	48.41
Media sul totale.....:	45.83	41.52	48.41
Scarto.....:	38.78	37.59	39.05

Fattore 3:	Totale	Maschi	Femmine
Totale.....:	53319	19228	31183
Rispondenti.....:	974	411	516
Media sui rispondenti:	54.74	46.78	60.43
Media sul totale.....:	54.74	46.78	60.43
Scarto.....:	36.64	37.44	34.85