

Salvatore Ingrassia
Carmela Senatore

Laboratorio di Statistica I

Guida alle Attività

Facoltà di Economia, Università della Calabria
Corso di Laurea in Statistica
Anno Accademico 2002-2003

Indice

1	L'Analisi dei dati con SPSS	1
1.1	Introduzione	1
1.2	Importazione di un file di dati	4
1.3	Preparazione di un file dati	9
1.3.1	Proprietà delle variabili	12
1.4	“German Credit Database”: descrizione dei dati	14
1.5	Sintassi	17
1.5.1	<i>Data List</i>	18
1.5.2	<i>Variables Lables</i>	19
1.5.3	<i>Values Labels</i>	20
1.5.4	<i>Missing Value</i>	20
1.5.5	<i>Informazioni sul file</i>	20
1.6	Esercizi	21

Capitolo 1

L'Analisi dei dati con SPSS

1.1 Introduzione

Chi si accinge a fare ricerca si imbatte subito in alcuni problemi di natura statistica che richiedono un certo numero di calcoli. Spesso questi calcoli sono semplici, ma ciò che è oneroso è la loro quantità; altre volte si necessita di una sola analisi ma di notevole complessità per il numero di variabili utilizzate. *Spss* è l'acronimo di *Statistical Package for the Social Sciences* ovvero *Pacchetto statistico per le scienze sociali* ed in effetti, l'SPSS si è affermato nel campo della ricerca sociale proprio per la sua filosofia di fondo.

Uno dei vantaggi, o dei difetti, dell'SPSS è quello di prendere alcune decisioni per conto del ricercatore che in questo modo non deve necessariamente conoscere tutti i passaggi matematici e statistici delle procedure che utilizza, ma può semplicemente come si interpretano i risultati ottenuti.

Un altro dei vantaggi è quello di essere un programma basato su un pseudo-linguaggio ovvero su una serie di parole chiave che servono per esprimere un comando caratterizzato da una sintassi e da una logica grammaticale; tutto ciò permette di avere una notevole elasticità nel tipo di calcoli da effettuare e nella scelta dei risultati che si vogliono ottenere.

Ovviamente un programma basato su menù a scelta multipla guidata è più facile da utilizzare anche se, non si conoscono interamente le possibilità di un programma; mentre un programma che utilizza un linguaggio, richiede la conoscenza iniziale dei termini fondamentali del linguaggio stesso.

SPSS è un programma studiato per soddisfare le esigenze di elaborazioni dati nei seguenti settori:

- sociologia e scienza politica;
- analisi/ricerche di mercato;
- medicina/biologia;

esso può eseguire le elaborazioni dati riguardanti le seguenti tecniche di analisi:

- analisi monovariata;
- analisi bivariata;
- analisi multivariata;
- test statistici parametrici e non parametrici;

- analisi di serie temporali;
- funzioni di sopravvivenza;
- modelli lineari;

e molte altre. Spss è uno dei più diffusi software per l'elaborazione statistica dei dati poichè permette di gestire i dati in un ambiente grafico; esso utilizza menù descrittivi e semplici finestre di dialogo per la maggior parte delle procedure. In SPSS ci sono diverse finestre¹ di dialogo:

- **Editor dei dati:** questa finestra visualizza il contenuto del file di dati. Possono essere creati nuovi file di dati o modificarne già esistenti. La finestra Data Editor si apre automaticamente con l'avvio di SPSS e può visualizzare solo un set di dati alla volta;
- **Viewer:** questa finestra visualizza tutti i risultati dell'analisi e può essere salvata per un successivo utilizzo. La visualizzazione di questa finestra avviene in modo automatico ogni qualvolta si esegue la procedura che genera l'*out put*.
- **Bozza:** questa finestra visualizza l'*out put* come semplice testo invece che come tabelle pivot interattive;
- **Editor di Tabelle Pivot:** questa finestra permette di modificare la visualizzazione dell'*out put*; ad esempio permette di modificare il posizionamento delle variabili nella tabella (colonne, righe..);
- **Editor dei grafici:** questa finestra permette di modificare colori, dimensioni, caratteri, gli assi di un grafico, far ruotare i grafici 3-D e3 anche cambiare il tipo di tabella.
- **Editor di output testuale:** questa finestra permette di modificare quella parte di testo relativa all'*out put* non visualizzata nella finestra pivot table editor.
- **Editor della sintassi:** questa finestra visualizza in forma di comandi di sintassi tutti le procedure che possono essere selezionate nelle finestre di dialogo durante l'analisi dei dati. In questo modo è possibile utilizzare particolari caratteristiche di SPSS non ottenibili attraverso le finestre di dialogo; ovviamente questi comandi possono essere salvati ed utilizzati per successive analisi.
- **Editor degli script:** questa finestra permette di personalizzare e automatizzare molte procedure in SPSS.

I dati utilizzati con Spss possono derivare da varie fonti: questionari, dati già raccolti, statistiche ufficiali; esso importa file provenienti anche da altri programmi quali ad esempio *Excel* o *Dbase*. Ogni matrice dati, di per sè non significa nulla, se non comunichiamo al programma di che tipo di dati si tratta, come sono organizzati nella matrice. Questo lavoro viene fatto costruendo un file, il DATA LIST, che tramite la sintassi di SPSS comunica al software tutte le informazioni necessarie per

¹Le funzionalità dei menù delle finestre variano in ogni finestra.

l'elaborazione e la comprensione dei risultati². Nella figura seguente sono illustrati i passaggi logici per arrivare dalla matrice di dati al system file SPSS:

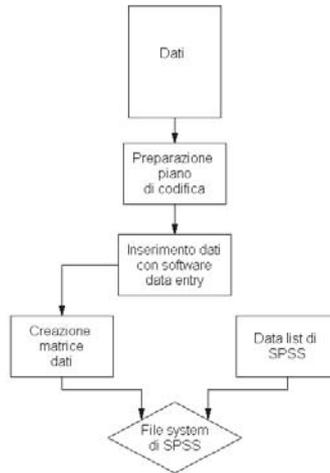


Figura 1.1: *Passaggi logici*

Creato il file di sistema di SPSS, si lavora sempre con lo stesso per svolgere le opportune elaborazioni statistiche. Il programma si apre con la seguente schermata che chiede che cosa si vuole fare:

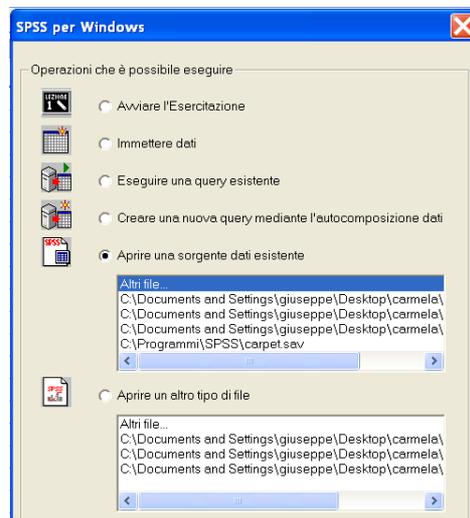


Figura 1.2: *Schermata Iniziale*

si può aprire una sorgente di file già esistente o aprire un altro tipo di file e quindi eventualmente procedere con l'importazione guidata di un file.

²Questa procedura può essere eseguita anche in modo automatico senza la visualizzazione della sintassi (vedi paragrafo 1.2).

Nel caso in cui si seleziona la voce annulla per aprire un file sarà necessario selezionare dal menù file la voce apri per accedere ad un file già esistente, od eventualmente la voce nuovo per generare un nuovo file (vedi Figura 1.3) in ogni caso le finestre che si possono aprire sono le seguenti:

Nome della finestra	Contenuto	Estensione del file
DATA	il system file spss	.sav
SYNTAX	Sintassi SPSS, DATA LIST, comandi	.sps
OUTPUT	Output dei comandi	.lst

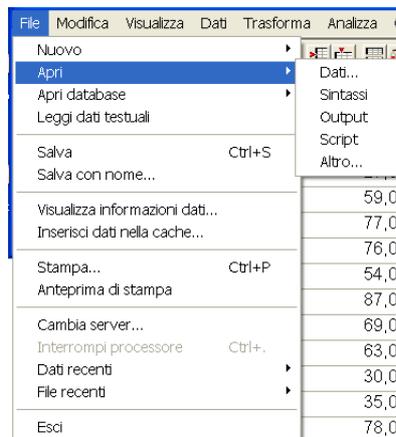


Figura 1.3: Menù

1.2 Importazione di un file di dati

La procedura da seguire per la creazione del file system è la seguente dal menù file si selezioni la voce apri quindi la voce dati si accede così alle finestre per l'importazione guidata del file. Questa procedura prevede sei passi qui di seguito descritti:

- nel primo passo il file viene visualizzato in una finestra di anteprima; è possibile scegliere di applicare un formato predefinito³ o eseguire i passi della procedura guidata per definire il metodo di lettura dei dati (vedi Figura 1.4);

³Il formato devessere stato salvato precedentemente.

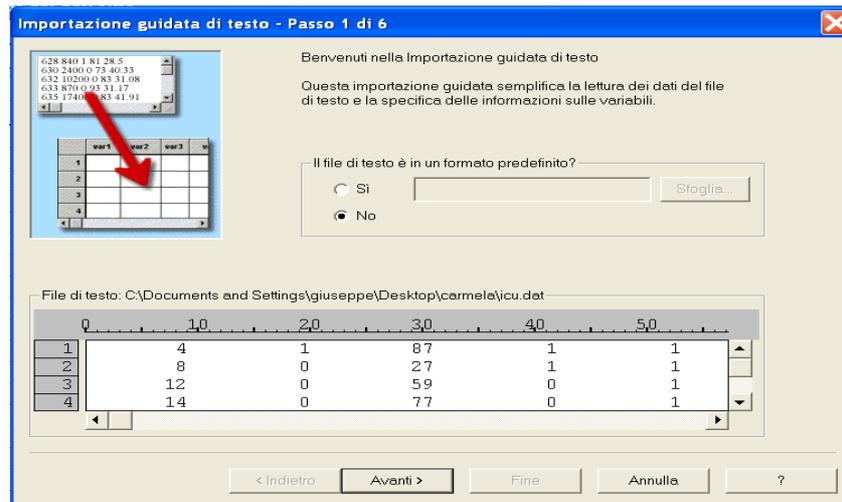


Figura 1.4: *Importazione guidata di testo: passo 1*

- nel secondo passo si devono fornire le informazioni sulle variabili. Per leggere correttamente i dati, durante l'importazione guidata si deve specificare dove finiscono i valori dei dati di una variabile e iniziano i valori di quella successiva. La disposizione delle variabili determina il metodo utilizzato per separare una variabile dalla successiva). I dati possono essere:
 - **Delimitati**: separati da spazi, virgole, tabulatori o altri caratteri . Le variabili di ciascun caso vengono registrate nello stesso ordine, ma non sempre nelle stesse posizioni di colonna;
 - **Larghezza fissa**: ciascuna variabile viene registrata nella stessa posizione della colonna e nella stesso record (riga) per ciascun caso del file di dati. Non è richiesto nessun delimitatore tra le variabili. In molti file di dati di testo generati automaticamente potrebbe sembrare che i valori dei dati siano mischiati insieme senza alcun simbolo di divisione; la posizione della colonna determina la variabile che verrà letta;

Inoltre è necessario specificare se la prima riga del file di dati contiene etichette descrittive che si vogliono utilizzare come nomi di variabili. Le etichette che includono più di otto caratteri verranno troncate. Se i primi otto caratteri non formano un nome univoco di variabile, il nome della variabile verrà modificato come necessario (vedi Figura 1.5);

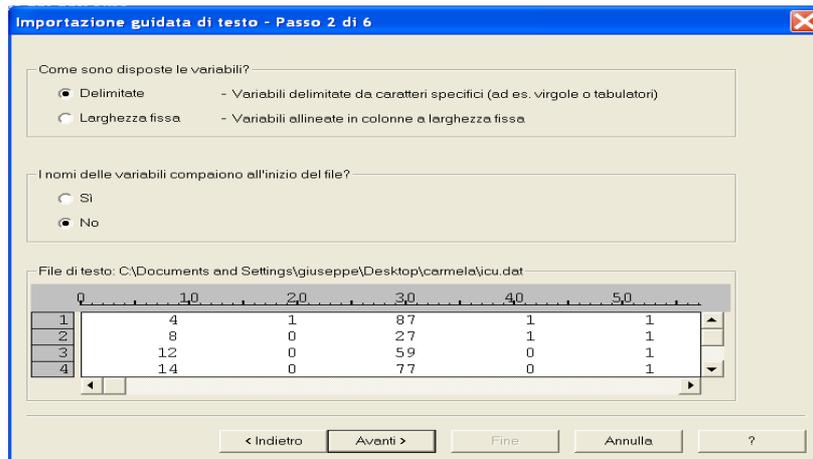


Figura 1.5: Importazione guidata di testo:passo 2

- nel terzo passo devono essere fornite informazioni sui casi a, in particolare si devono specificare le seguenti voci (vedi Figura 1.6):

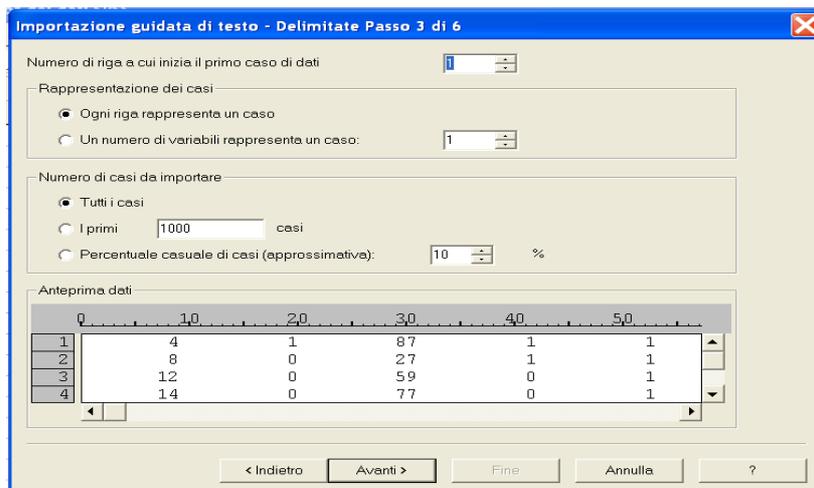


Figura 1.6: Importazione guidata di testo:passo 3

- **Numero di riga a cui inizia il primo caso dati:** indica la prima riga del file di dati che contiene valori. Se le prime righe del file di dati contengono etichette descrittive o testo di altro tipo che non rappresenta un valore, questa non sarà la riga 1;
- **Numero di righe che rappresentano un caso:** nel caso di file a larghezza fissa controlla il modo in cui si determina ogni caso e l'inizio del successivo. Ogni riga è definita dal relativo numero di riga all'interno del caso e della posizione della colonna; è necessario specificare il numero di righe per ogni caso per poter leggere correttamente i dati;
- **Numero di casi da importare:** è possibile importare tutti i casi del file di dati, solo i primi n casi o un campione casuale determinato in base a una percentuale specificata.

Poichè viene eseguito un processo indipendente di decisione pseudo-casuale per ciascun caso, l'equivalenza tra la percentuale di casi selezionati e la percentuale specificata può essere solo approssimativa. Maggiore è il numero di casi inclusi nel data file e maggiore sarà l'approssimazione della percentuale di casi selezionati rispetto alla percentuale specificata;

- nel quarto passo, è illustrata la soluzione più opportuna per leggere un file di dati e consente di modificare il metodo impiegato per leggere le variabili dal file di dati. Quando il file è a larghezza fissa le linee verticali nella finestra di anteprima indicano le posizioni in cui la procedura ipotizza possa iniziare una variabile. Si possono inserire, spostare ed eliminare linee di interruzione delle variabili in base alle necessità. Quando il file è delimitato è necessario specificare i delimitatori tra le variabili (spazi, virgole,...) e se sono presenti più delimitatori consecutivi, questi sono considerati valori mancanti; inoltre è necessario specificare il qualificatore di testo (vedi Figura 1.7);

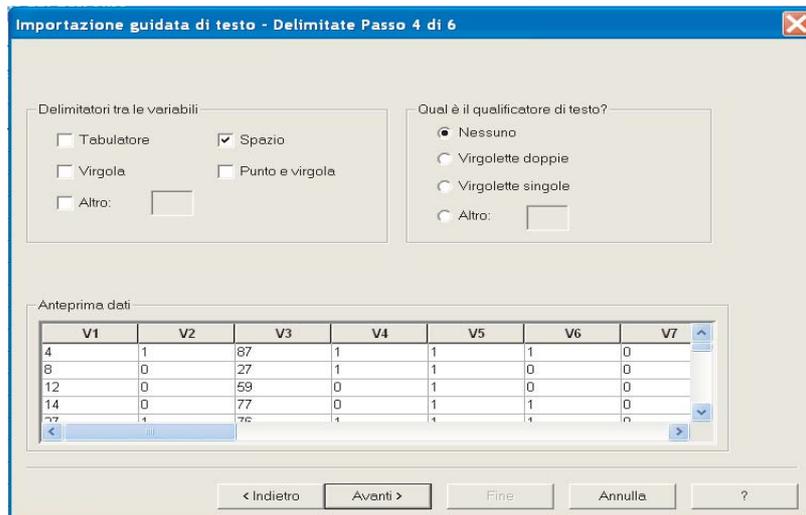


Figura 1.7: *Importazione guidata di testo: passo 4*

- nel quinto passo si controlla il nome della variabile e il formato dei dati usati in modo da leggere le variabili e scegliere quelle che saranno incluse nel file di dati finale (vedi figura 1.8);

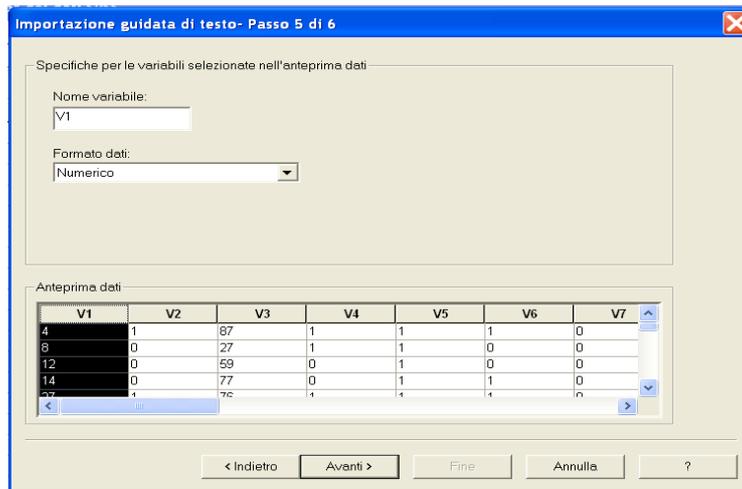


Figura 1.8: Importazione guidata di testo: passo 5

- nel sesto ed ultimo passo è possibile salvare le impostazioni in un file da utilizzare per l'importazione di file di dati di testo simili. Inoltre si può incollare la sintassi generata e quindi poterla personalizzare e salvare in modo da poterle utilizzare in altre sessioni o durante i processi di produzione (vedi figura 1.9).

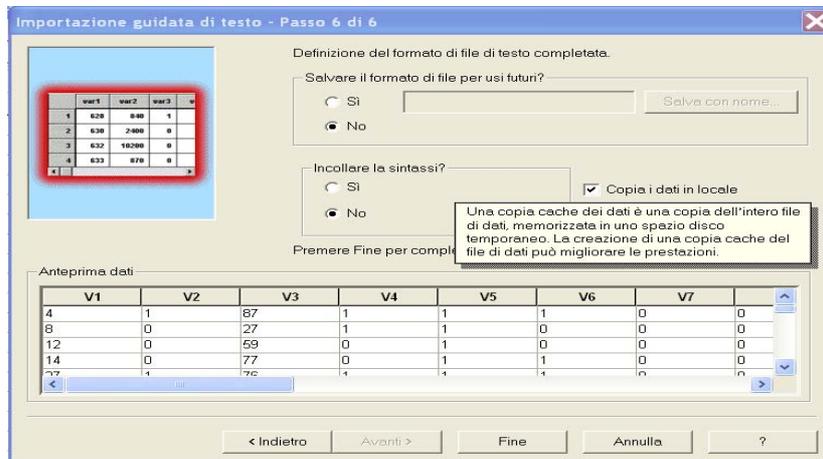


Figura 1.9: Importazione guidata di testo: passo 6

Completate le operazioni di esportazione il file di dati si presenterà come in figura 1.10.

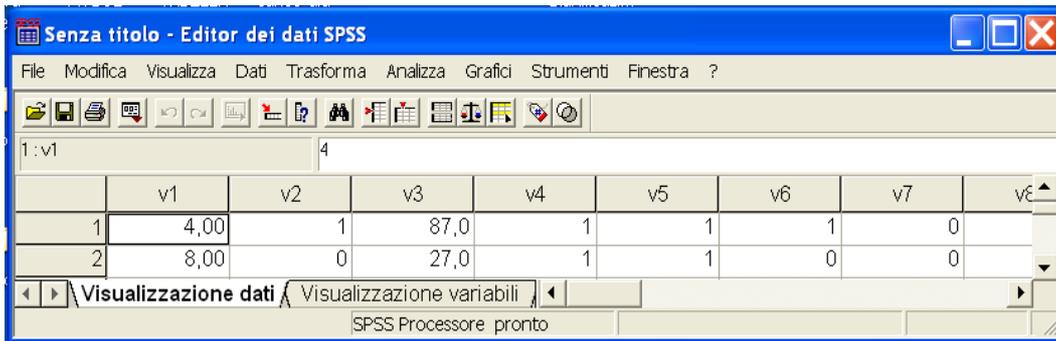


Figura 1.10: *Editor dei dati.*

1.3 Preparazione di un file dati

Come si può notare dalla figura 1.10, l'editor dei dati, è un file costituito da due fogli in uno sono visualizzati i dati e nell'altro le variabili (vedi figura 1.11).

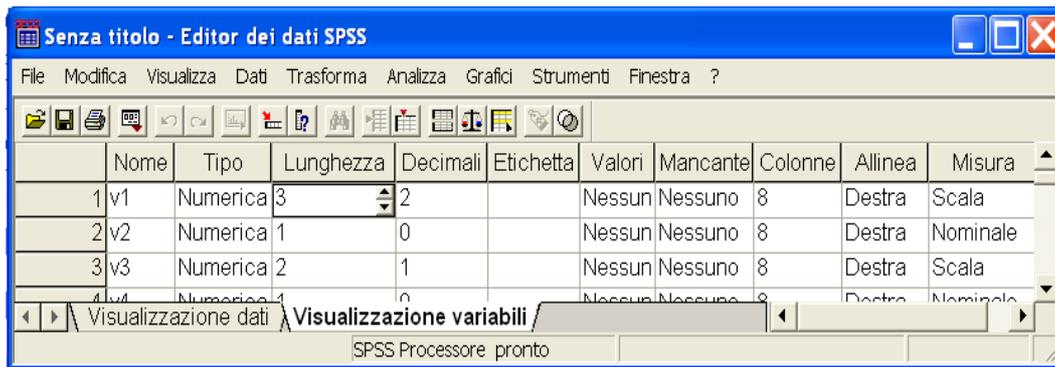


Figura 1.11: *Editor delle variabili.*

Il foglio relativo alle variabili è costituito da dieci campi che possono essere impostati per ogni variabile, ogni record corrisponde ad una variabile. Gli attributi da definire per ogni variabile coinvolta nell'analisi sono i seguenti:

- nome: in questo campo deve essere inserito il nome della variabile se non è stato già inserito durante l'importazione guidata del testo. I nomi non devono superare gli otto caratteri, non possono contenere spazi bianchi e ne terminare con il come pure con caratteri speciali quali (! ? " *); inoltre possono essere usati una sola volta, cioè non è possibile attribuire lo stesso nome a due variabili diverse;
- tipo: in questo campo si specifica il tipo di variabile numerica, data... La selezione di questa caratteristica avviene attraverso la finestra di dialogo Tipo di variabile che è automaticamente aperta quando si clicca sulla cella (vedi Figure 1.12);



Figura 1.12: *Tipo di variabile.*

inoltre da questa finestra si può cambiare la larghezza standard di otto colonne in quella sufficiente alla descrizione dei dati reali e decidere il numero di decimali per le variabili numeriche;

- lunghezza: definisce la lunghezza della variabile
- decimali: in questo campo si definiscono innumeri decimali per le variabili numeriche;
- etichetta: in questo campo si assegnano le etichette alle variabili, se ne da cioè una descrizione;
- valori: in questo campo si possono assegnare valori descrittivi per ogni valore assunto da una variabile. Questo campo risulta particolarmente utile quando si lavora con file con codici numerici per variabili non numeriche; per accedere alla finestra di dialogo Etichette dei dati si procede come nel caso del campo tipo (vedi Figura 1.13);



Figura 1.13: *Etichette dei dati.*

nel riquadro valore si digita il codice e nella casella etichetta di valore la modalità, cliccando il pulsante aggiungi si vedrà comparire quanto precedentemente digitato; ripetendo l'operazione, si possono aggiungere altre etichette e valori.

- **mancante**: in questo campo si possono definire specificati valori dei dati come valori mancanti; nella finestra di dialogo dati mancanti (vedi Figura 1.14), alla quale si accede come nelle finestre precedentemente descritte, si possono definire differenti valori mancanti per ogni variabile, oppure si può definire un intervallo di valori discreti all'interno del quale tutti i valori vengono considerati mancanti;



Figura 1.14: *Dati mancanti.*

- **colonne**: in questo campo si specifica il numero di caratteri per la larghezza della colonna;
- **allinea**: in questo campo si imposta l'allineamento delle variabili nel foglio dei dati, di default l'allineamento è a destra per i valori numeri e a sinistra per le stringhe;
- **misura**: in questo campo si definisce la scala della variabile che si seleziona cliccando sulla freccia verso il basso che è visualizzata nel momento che si seleziona la cella di interesse (vedi Figura 1.15);



Figura 1.15: *Misura della variabile.*

i livelli di misurazione che possono essere scelti sono i seguenti:

- **nominale**, che si seleziona quando fra le modalità del carattere non è possibile istituire nessun ordinamento;
- **ordinale**, che si seleziona quando è possibile istituire un ordine naturale fra le modalità del carattere;
- **scala**, che si seleziona quando i numeri cardinali sono impiegati per indicare delle quantità.

La definizione del livello di misurazione è fondamentale per l'analisi, poichè al suo variare varia il tipo di analisi che può essere condotta; ad esempio non ha senso logico istituire dei rapporti fra le intensità di caratteri nominali.

1.3.1 Proprietà delle variabili

La definizione delle proprietà delle variabili può essere fatta, non solo dal foglio di visualizzazione delle variabili, ma anche selezionando dal menù dati la voce *Definisci Proprietà Variabili*. La selezione di questa voce farà visualizzare la finestra di dialogo riportata in figura 1.16:

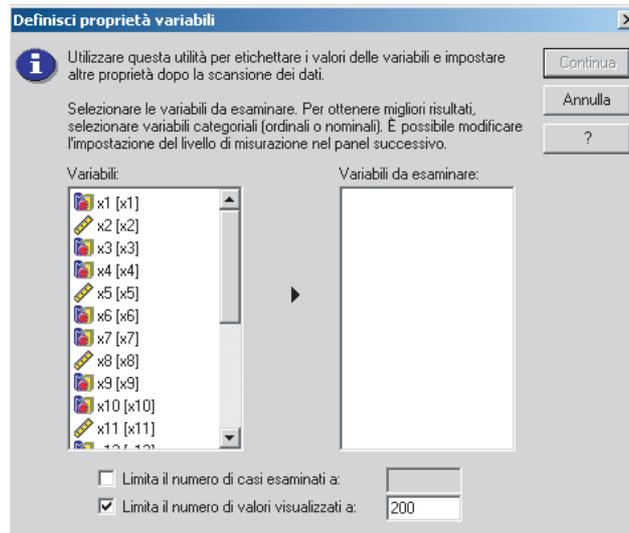


Figura 1.16: *Proprietà delle variabili*.

in questa prima finestra dovranno essere selezionate le variabili di cui si vogliono definire le caratteristiche ed eventualmente specificare il numero di di casi da esaminare, fatta la scelta nel passo successivo sarà visualizzata la finestra riportata in figura 1.17.

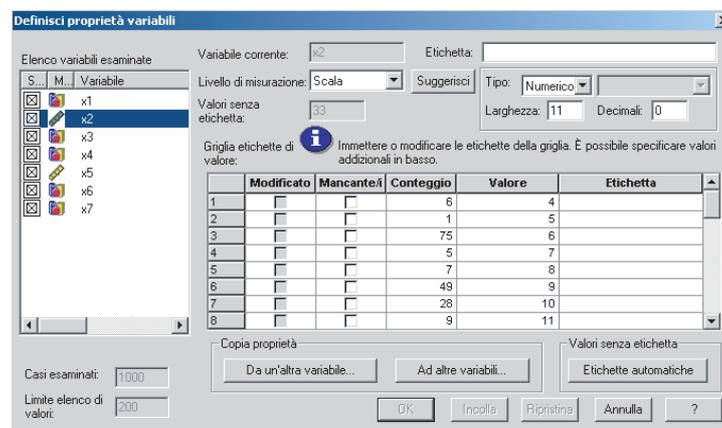


Figura 1.17: *Proprietà delle variabili*.

In quest'ultima finestra, dopo aver selezionato una variabile, si potranno creare etichette di valori o definire/modificare altre proprietà della variabile (per la descrizione delle proprietà vedere paragrafo 1.2). L'operazione di definizione delle caratteristiche deve essere ripetuta per ogni singola variabile.

Infine un'ultima procedura per la definizione delle proprietà delle variabili è quella della copia delle proprietà dei dati da file già organizzati. La funzione *copia proprietà dei dati*, alla quale si accede dal menù dati, permette di copiare etichette di valori e altre proprietà di variabili; la procedura per la copia è costituita di cinque passi:

- nel primo passo si sceglie l'origine delle proprietà che può essere un file SPSS esterno o lo stesso file di lavoro (vedi figura 1.18);
- nel secondo passo si selezionano le variabili di cui si vogliono copiare le proprietà e le variabili che si vogliono modificare, le prime dall'elenco sorgente le seconde dall'elenco destinazione (vedi figura 1.18);

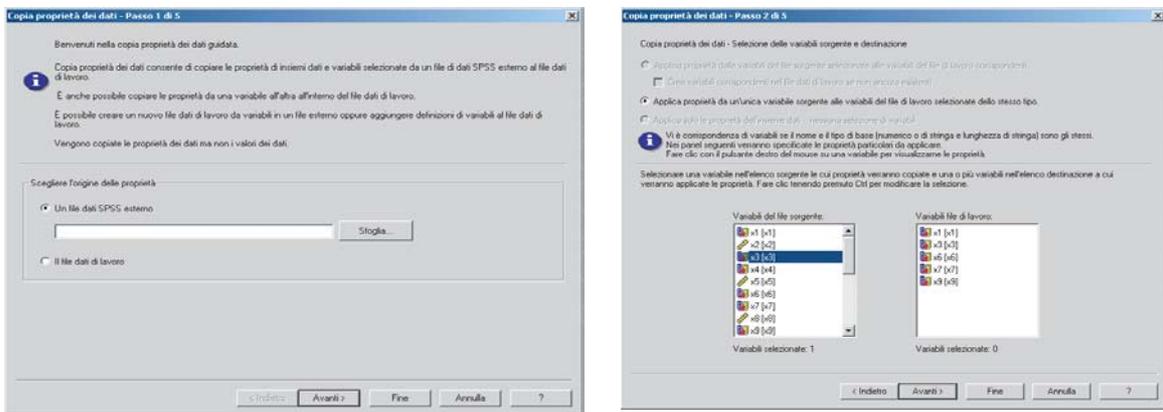


Figura 1.18: Copia proprietà delle variabili.

- nel terzo passo si selezionano le proprietà che si vogliono copiare per le variabili selezionate esistenti (vedi figura 1.19);
- nel quarto passo le proprietà dell'insieme di dati d'origine (vedi figura 1.19);

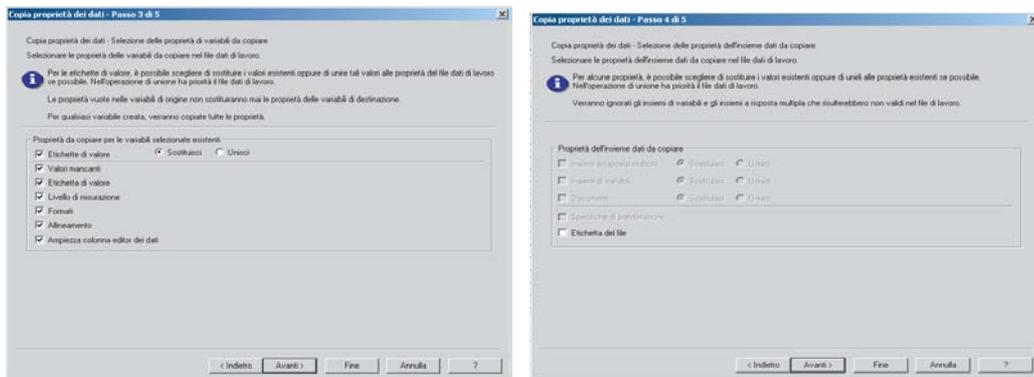


Figura 1.19: Copia proprietà delle variabili.

- nel quinto ed ultimo passo, sono sintetizzate le modifiche apportate alle variabili e si può scegliere se eseguire direttamente il comando o incollarlo nella finestra di sintassi (vedi figura 1.20);

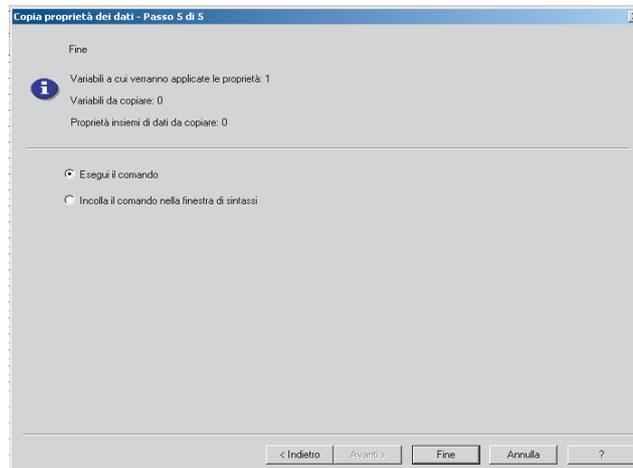


Figura 1.20: Copia proprietà delle variabili.

1.4 “German Credit Database”: descrizione dei dati

Si consideri il set di dati qui di seguito descritto e si importi e prepari il file system di SPSS con le procedure appena descritte. Il file di dati german.dat è composto da 1000 osservazioni che individuano soggetti che hanno richiesto un credito; 700 individui si sono rivelati solventi buoni pagatori, mentre 300 si sono rivelati insolventi cattivi pagatori. I primi costituiscono il gruppo “good” contrassegnato con 1, i secondi il gruppo “bad” contrassegnato con 2. I crediti in questione sono di piccolo importo (importo massimo circa pari a 16000 DM) e breve durata (durata massima 60 mesi). Le informazioni relative ai richiedenti credito sono descritte da 20 variabili (attributi), 7 continue e 13 categoriche.

Le variabili sono così definite:

- Variabile X_1 : Saldo attuale del conto corrente, variabile categorica ordinale a 4 modalità:
 - A_{11} : $X_1 < 0DM$
 - A_{12} : $0 \leq X_1 < 200DM$
 - A_{13} : $X_1 \geq 200DM$ / salario ricevuto per almeno un anno
 - A_{14} : nessun conto corrente
- Variabile X_2 : Durata del prestito in mesi, variabile quantitativa
- Variabile X_3 : Storia del credito, variabile categorica nominale a 5 modalità
 - A_{30} : nessun credito richiesto/tutti i crediti pagati in tempo
 - A_{31} : tutti i crediti richiesti a questa banca pagati in tempo

- A_{32} : crediti attuali fin ora pagati in tempo
 - A_{33} : ritardi nei pagamenti in impegni passati
 - A_{34} : conto critico/altri crediti esistenti (non in questa banca)
 - Variabile X_4 : Scopo della richiesta di credito, variabile categorica nominale a 11 modalità
 - A_{40} : auto (nuova)
 - A_{41} : auto (usata)
 - A_{42} : mobilio/attrezzature
 - A_{43} : radio/televisione
 - A_{44} : elettrodomestici
 - A_{45} : riparazioni
 - A_{46} : istruzione
 - A_{47} : vacanza
 - A_{48} : riqualificazione
 - A_{49} : affari
 - A_{410} : altro
 - Variabile X_5 : Ammontare del credito richiesto, variabile quantitativa
 - Variabile X_6 : Tipologia conto di risparmio/obbligazioni, variabile categorica ordinale a 5 modalità
 - A_{61} : $X_6 < 100DM$
 - A_{62} : $100 \leq X_6 < 500DM$
 - A_{63} : $500 \leq X_6 < 1000DM$
 - A_{64} : $X_6 \geq 1000DM$
 - A_{65} : nessun conto di risparmio/sconosciuto
 - Variabile X_7 : Anni di assunzione dell'attuale impiego, variabile categorica ordinale a 5 modalità
 - A_{71} : disoccupato
 - A_{72} : $X_7 < 1$ anno
 - A_{73} : $1 \leq X_7 < 4$ anni
 - A_{74} : $4 \leq X_7 \leq 7$ anni
 - A_{75} : $X_7 \geq 7$ anni
 - Variabile X_8 : Stima della rata in percentuale all'entrata disponibile, variabile quantitativa
 - Variabile X_9 : Stato civile e sesso, variabile categorica nominale a 5 modalità
-

- A_{91} : maschio: divorziato/separato
- A_{92} : femmina: divorziata/separata/sposata
- A_{93} : maschio: single
- A_{94} : maschio: sposato/vedovo
- A_{95} : femmina: single
- Variabile X_{10} : Altri debitori/garanti, variabile categorica nominale a 3 modalità
 - A_{101} : nessuno
 - A_{102} : co-richiedente
 - A_{103} : garante
- Variabile X_{11} : Anni dell'attuale residenza, variabile quantitativa
- Variabile X_{12} : Proprietà, variabile categorica nominale, 4 modalità
 - A_{121} : beni immobili
 - A_{122} : se non è A_{121} : società di credito edilizio/assicurazione sulla vita
 - A_{123} : se non è A_{121} o A_{122} : auto o altro, non presente nella variabile 6
 - A_{124} : sconosciute/nessuna proprietà
- Variabile X_{13} : Età, variabile quantitativa
- Variabile X_{14} : Altri schemi di pagamento, variabile categorica nominale a 3 modalità
 - A_{141} : banca
 - A_{142} : negozi
 - A_{143} : nessuno
- Variabile X_{15} : Casa di abitazione, variabile categorica a tre modalità
 - A_{151} : affitto
 - A_{152} : di proprietà
 - A_{153} : ospite
- Variabile X_{16} : Numero di crediti attualmente in questa banca, variabile quantitativa
- Variabile X_{17} : Lavoro, variabile categorica nominale a 4 modalità
 - A_{171} : disoccupato/non-specializzato, non-residente
 - A_{172} : non specializzato residente
 - A_{173} : impiegato specializzato/funzionario
 - A_{174} : dirigente/autonomo/impiego altamente qualificato/ufficiale

- Variabile X_{18} : Numero di persone a carico, variabile quantitativa
- Variabile X_{19} : Telefono, variabile categorica nominale a 2 modalità
 - A_{191} : no
 - A_{192} : si
- Variabile X_{20} : Lavoratore straniero, variabile categorica nominale a 2 modalità
 - A_{201} : si
 - A_{202} : no.

1.5 Sintassi

Nel paragrafo 1.2 si è visto come contemporaneamente alla finestra di editor dei dati, attraverso la copia della sintassi, si può ottenere il file di sintassi; in figura 1.21 si osserva una finestra di editor di sintassi.

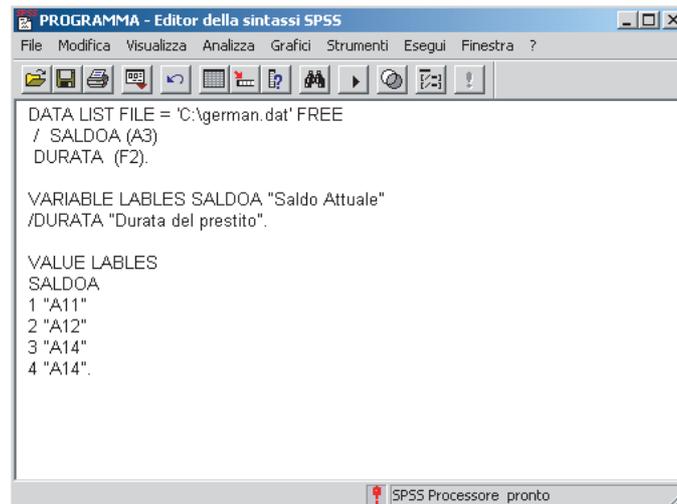


Figura 1.21: *Editor di sintassi.*

In SPSS sono presenti vari comandi e funzioni che possono essere suddivisi in tre principali categorie (vedi tabella):

1. **Comandi di definizione dei dati**, che forniscono informazioni sulle variabili, il tipo e la loro locazione nel file di dati;
2. **Comandi di trasformazione dei dati**, che sono usati per creare nuove variabili, modificare quelle esistenti e selezionare un sottocampione di casi per l'analisi;
3. **Comandi che attivano procedure**, come le statistiche, i grafici, il salvataggio dei dati.

Categoria	Comandi
Definizione	data list, variable labels, value labels, missing value
Trasformazione	compute, recode, if, temporary, select if
Procedure	Frequencies, Crosstabs, Save outfile

Un'istruzione spss è sempre composta da un comando principale seguito da altre parole chiavi e completato con informazioni fornite dall'utente. Per eseguire i comandi di un file basat selezionare gli stessi e cliccare sul tasto esegui corrente posizionato sulla barra degli strumenti. Alcune volte può capitare che durante l'esecuzione della procedura il processore mantenga delle trasformazioni in sospeso, per evitare basta scegliere dal menù modifica la voce opzioni, quindi nella scheda dati selezionare la voce *esegui le trasformazioni immediatamente* (vedi figura 1.22).



Figura 1.22: Esecuzione comandi.

In questo capitolo saranno trattati i principali comandi per la definizione di dati in modo da poter generare e preparare il file system di spss anche dall'editor della sintassi. Questa operazione richiede:

- l'impostazione del **DATAT LIST**, il comando, cioè che dice al programma dove i dati grezzi possono essere trovati e definisce per ciascuna variabile il nome e il numero iniziale e finale delle colonne occupate. Il **DATA LIST** è indispensabile: senza di esso il programma non riesce a distinguere tra loro le diverse variabili;
- la definizione delle **VARIABLES LABELS**, il comando, cioè, che permette di attribuire a ciascuna variabile un'etichetta esplicativa tale da permettere una migliore leggibilità dei risultati delle elaborazioni.
- la definizione delle **VALUE LABELS**, il comando che fornisce etichette descrittive per i valori delle variabili.

1.5.1 Data List

Esistono due tipi di **DATA LIST**: quello per file di dati in formato fisso e quello per file di dati in formato delimitato.

- *Data List Fixed* si usa quando i valori di ogni variabile sono registrati, nel file di dati grezzo, nella stessa posizione per ogni caso;

- *Data List Free* si usa quando le variabili sono registrate, nel file di dati grezzo, nello stesso ordine per ogni caso, ma non necessariamente nella stessa posizione; in questo caso il data list si semplifica di molto poichè non è necessario scrivere la colonna di inizio e di fine occupata dai valori di ogni variabile.

Per una corretta esecuzione il comando deve essere scritto nel seguente ordine: nella prima riga deve essere digitato il nome del comando seguito dal *path* del file che contiene i dati grezzi e dal parametro FIXED o FREE in funzione del di formato dei dati grezzi. Nelle righe successive si assegnano i nomi alle variabili, il numero iniziale e finale delle colonne occupate ed il tipo di variabile. Più precisamente devono essere rispettate le seguenti condizioni:

- al nome delle variabile numeriche, che contengono cioè solo numeri, segue l'indicazione delle colonne occupate (se le colonne sono 2 o più è sufficiente indicare la prima e l'ultima separate da un trattino,-) ed eventualmente il numero di decimali tra parentesi;
- le variabili alfanumeriche o di tipo stringa, che possono contenere una combinazione di lettere e numeri, e altri caratteri sono contraddistinte da una (A) posta dopo l'indicazione seguita dal numero di caratteri che la costituiscono; ad esempio nella caso del german data set la prima variabile è del tipo (A3). Nel caso di variabili numeriche il simbolo è la lettera F seguita da due numeri seguiti da un punto, il primo indica il numero dei valori interi il secondo il numero dei decimali; ad esempio se una variabile assume valori inferiori a 100 e con al massimo 2 cifre decimali la giusta definizione sarebbe data da (F2.2);
- il punto, ., alla fine di questo set di istruzioni è obbligatorio.

Ovviamente nel definire le variabili devono essere rispettate le stesse condizioni di cui al paragrafo 1.3, in figura 1.21 si osserva un esempio di data list.

1.5.2 Variables Labels

Questo comando assegna un'etichetta descrittiva (label) ad ogni variabile. Le etichette possono comprendere spazi bianchi e qualsiasi altro carattere, ogni etichetta può raggiungere fino a 256 caratteri. Anche in questo caso il nome del comando deve essere digitato nella prima riga e poi seguito dai nomi delle variabili e le relative descrizioni, le uniche accortezze che devono essere poste sono le seguenti:

- i nomi delle variabili devono coincidere con quelli del DATA LIST;
- ogni variabile va separata dalla successiva con uno slash;
- un'etichetta non pu essere spezzata in pi righe;
- l'inizio e la fine delle label deve essere segnalata con le virgolette.

In figura 1.21 si osserva un esempio di variables labels.

1.5.3 Values Labels

Dopo aver definito la lista delle variabili con il DATA LIST e le etichette (label) delle variabili con VARIABLE LABEL non resta che attribuire delle etichette a ciascun valore di ogni variabile con il comando VALUE LABELS. Anche in questo caso il nome del comando deve essere digitato nella prima riga e poi seguito dai valori da attribuire alle relative variabili, le uniche accortezze che devono essere poste sono le seguenti:

- ogni value labels non deve superare i 60 caratteri;
- il nome delle variabile indicata deve coincidere con quello delle variabel labels e del data list;
- ogni variabile va separata dalla successiva con uno slash /;
- l'inizio e la fine delle label segnalata con le virgolette.

In figura 1.21 si osserva un esempio di value lables.

1.5.4 Missing Value

Qualche volta l'informazione per un particolare caso non è utilizzabile si pu allora assegnare, con il comando MISSING VALUE, valori per identificare le mancate risposte. Ad esempio con il comando: MISSING VALUE PROVA (0) attribuisce al valore 0 della variabile PROVA il significato di valore mancante per l'utente.

1.5.5 Informazioni sul file

Qualsiasi sia le procedura utilizzata per ottenere il file system, sul quale condurre le elaborazioni, al termine della procedura SPSS offre sempre la possibilità di visualizzare un file di tipo output, nel quale sono descritte le caratteristiche definite per ogni variabile del data set. A questo file si accede dal menù strunenti dall'editor dei dati (vedi Figura 1.23).

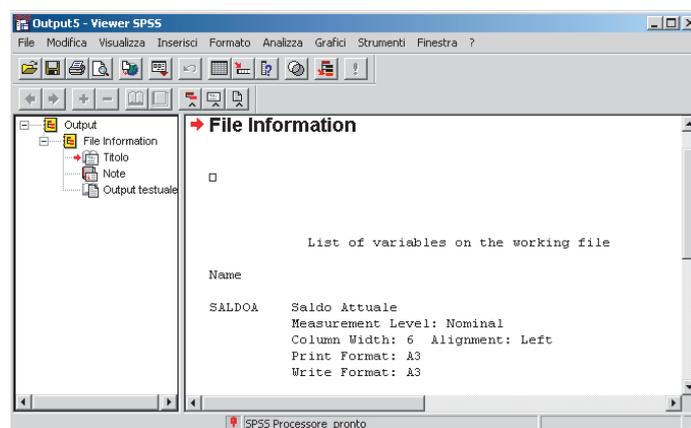


Figura 1.23: Informazione sul file di dati.

1.6 Esercizi

Esercizio 1.1

Importare i file `icu.data` e `icu.txt` reperibili dal sito

www.economia.unical.it/STATistica/Laboratori/dati/wine.dat e costruire il file `systemm` di `spss` secondo le procedure previste dalle finestre di dialogo e poi attraverso i comandi di sintassi.