



La cassetta degli attrezzi. Strumenti per le scienze umane

direttore

Giovanni Di Franco, Università di Salerno

comitato editoriale

Elena Battaglini, Ires-Cgil;  
Sara Bentivegna, Università di Roma;  
Alberto Marradi, Università di Firenze;  
Federica Pintaldi, Istat;  
Luciana Quattrocioni, Istat;  
Marta Simoni, Iref-Acli.

La collana, rivolta a ricercatori accademici e professionisti, studiosi, studenti, e operatori del variegato mondo della ricerca empirica nelle scienze umane, si colloca sul versante dell'alta divulgazione e intende offrire strumenti di riflessione e di intervento per la ricerca.

Obiettivo è consolidare le discipline umane presentando gli strumenti di ricerca empirica, sia di raccolta sia di analisi dei dati, in modo intellegibile e metodologicamente critico così da consentirne l'applicazione proficua rispetto a definiti obiettivi cognitivi.

I testi sono scritti da professionisti della ricerca che, attingendo alla personale esperienza maturata in anni di attività, offrono ai lettori strumenti concettuali e tecnici immediatamente applicabili nella propria attività di ricerca.

Tutti i volumi pubblicati sono sottoposti a referaggio.



Giovanni Di Franco

# **L'ANALISI DEI DATI CON SPSS**

Guida alla programmazione  
e alla sintassi dei comandi

La cassetta degli attrezzi  
Strumenti per le scienze umane/I

**FrancoAngeli**

Progetto grafico di copertina di Maria Teresa Pizzetti

Copyright © 2009 by FrancoAngeli s.r.l., Milano, Italy.

*L'opera, comprese tutte le sue parti, è tutelata dalla legge sul diritto d'autore. L'Utente nel momento in cui effettua il download dell'opera accetta tutte le condizioni della licenza d'uso dell'opera previste e comunicate sul sito [www.francoangeli.it](http://www.francoangeli.it).*

119. *La cassetta degli attrezzi. Strumenti per le scienze umane*

1. Giovanni Di Franco, *L'analisi dei dati con Spss. Guida alla programmazione e alla sintassi dei comandi*
2. Silvia Cataldi, *Come si analizzano i focus group*
3. Federica Pintaldi, *Come si analizzano i dati territoriali*
4. Giovanni Di Franco, *Il campionamento nelle scienze umane: come ridurre la distanza fra la teoria e la pratica*
5. Alberto Marradi, *Come evitare gli errori tipici in un questionario*



# Indice

<b>1. Introduzione</b>	pag.	9
<b>2. Il programma Spss</b>	»	17
2.1 L'ambiente di lavoro Spss	»	19
2.2 Come costruire una matrice dei dati in Spss	»	30
2.3 Come importare una matrice dei dati da Excel	»	43
2.4 Le prime analisi con Spss	»	48
2.5 Cosa leggere per saperne di più	»	54
<b>3. La sintassi di Spss</b>	»	57
3.1 I primi comandi con la sintassi	»	59
3.2 Come si mandano in esecuzione i comandi dalla sintassi	»	65
3.3 Tipi di comandi	»	68
3.4 Regole generali della sintassi di Spss	»	73
3.5 Cosa leggere per saperne di più	»	75
<b>4. Definire, manipolare e trasformare i dati</b>	»	77
4.1 Comandi di definizione dei dati	»	78
4.2 Comandi di trasformazione delle variabili	»	82
4.3 Comandi di trasformazione della matrice dei dati	»	99
4.4 Comandi di selezione e ponderazione dei casi	»	105
4.5 Cosa leggere per saperne di più	»	110
<b>5. Le procedure di analisi dei dati</b>	»	111
5.1 Procedure descrittive	»	111
5.2 Procedure che comparano gruppi	»	119
5.3 Procedure di analisi della correlazione	»	124
5.4 Procedure di analisi della regressione	»	127

5.5	Procedure di analisi log-lineare	»	135
5.6	Procedure di classificazione automatica	»	143
5.7	Procedure di riduzione dimensionale dei dati	»	153
5.8	Procedure di scaling	»	163
5.9	Procedure per il trattamento delle domande a risposta multipla	»	168
5.10	Cosa leggere per saperne di più	»	171

## 1. Introduzione

Il programma Spss (Statistical Package for Social Science; vedi il sito <http://www.spss.it>) è dalla fine degli anni '60 del secolo scorso il più diffuso *software* di analisi dei dati presso i ricercatori delle scienze umane. Naturalmente, ciò non vuol dire che Spss sia l'unico o il migliore programma di analisi dei dati attualmente disponibile sul mercato; ma che esso conta più di due milioni di utenti in tutto il mondo.

A partire dalla versione 6.0 (nel 2009 è stata licenziata la versione 17.0) è compatibile con il sistema operativo Windows; le precedenti versioni erano compatibili con il sistema operativo Dos. Sono disponibili versioni di Spss anche per i sistemi operativi Mac e Linux.

Spss consente di applicare molte procedure di gestione e di manipolazione dei dati e di analisi statistica. Le procedure di analisi disponibili dipendono dal numero di moduli del programma di cui si dispone. Infatti, Spss è un programma modulare integrato. Ciascun modulo include alcune procedure di analisi statistica. Il modulo base di Spss è l'unico essenziale e del tutto autosufficiente. Esso consente le operazioni di gestione dati, di trasformazioni dei dati, la realizzazioni di grafici, le procedure statistiche monovariate, bivariate e multivariate. Di solito, la maggior parte delle esigenze di analisi degli utenti sono soddisfatte dalle procedure presenti nel modulo base. Gli altri moduli di Spss sono invece dedicati a procedure avanzate di analisi dei dati che soddisfano le esigenze di utenti specializzati in particolari settori di ricerca. Di seguito si elencano le caratteristiche principali di alcuni dei moduli aggiuntivi che possono integrare il modulo base del programma:

- *Regression Models*: diversi modelli di regressione lineare e non lineare;
- *Advanced Models*: modelli di analisi della varianza e/o covarianza; modelli GLM (modello lineare generalizzato); modelli log-lineari;
- *Tables*: diverse procedure per la gestione del formato di presentazione delle tabelle;
- *Trends*: procedure per l'analisi di serie storiche;
- *Categories*: procedure di analisi per variabili categoriali (scaling ottimale; analisi delle corrispondenze binarie; analisi delle corrispondenze multiple, etc.);
- *Conjoint*: procedure per l'analisi delle preferenze dei consumatori;
- *Maps*: mappe tematiche integrate, cartografia;
- *Exact Tests*: procedure statistiche per piccoli campioni;
- *Missing Values Analysis*: procedure per il trattamento dati mancanti;
- *Amos*: modelli Lisrel (modelli di equazioni strutturali);
- *TextSmart*: programmi per l'analisi di testi (statistica testuale);
- *Sample Power*: programma per determinare la numerosità dei campioni;
- *Data Entry*: programma di inserimento dati (specie dati *survey*);
- *Neural Connection*: programmi per la costruzione reti neurali artificiali.

Oltre a quelli elencati ci sono altri moduli che qui non abbiamo ritenuto opportuno presentare. I lettori interessati possono consultare il sito italiano ([www.spss.it](http://www.spss.it)).

È chiara l'impostazione generalista di Spss che si propone come un pacchetto di programmi che ciascun utente può comporre a suo piacimento soddisfacendo in modo completo tutte le sue esigenze. Come detto, per la maggior parte degli utenti tipo possono bastare e avanzare le procedure che sono incluse nel modulo base.

Nel presente volume prenderemo in considerazione soprattutto le procedure di pre-analisi e di analisi dei dati che sono comunemente usate dalla maggior parte dei ricercatori.

In un manuale è necessario compiere delle scelte per evitare di sovraesporre il lettore ad una mole eccessiva di informazioni che potrebbero indurre un effetto di confusione e scoraggiamento. Per evitare un tale effetto abbiamo delimitato i contenuti alle principali operazioni che di solito si svolgono nelle analisi dei dati.

Una volta acquisite le operazioni per così dire di *routine*, ciascun lettore, qualora fosse interessato, sarà in grado autonomamente di approfondire tutte le altre procedure di analisi che il programma offre. L'obiettivo prioritario sarà permettere ai lettori un uso professionale del programma e ciò comporta la necessità di fissare l'attenzione sulla sintassi dei comandi piuttosto che su un approccio più *friendly* costituito dalla modalità di lavoro attraverso i menù e le finestre di dialogo.

Prima dell'avvento del sistema operativo Windows, per poter impostare i comandi e ottenere i relativi risultati, gli utenti di Spss dovevano necessariamente conoscere la sintassi del programma. Conoscere la sintassi significa imparare un linguaggio di programmazione molto semplice – che non ha nulla a che fare con i decisamente più complessi linguaggi di programmazione informatici – e, quindi, essere in grado di dialogare, naturalmente in senso metaforico, con il programma. In quei tempi, ossia quando il sistema operativo era il Dos (Disk Operation System), tutti i programmi cosiddetti applicativi, e lo stesso sistema operativo funzionavano attraverso le rispettive sintassi dei comandi. Ad esempio, un utente medio che nel suo lavoro aveva bisogno di un programma di scrittura, di un foglio elettronico, di un programma di archiviazione dei dati e di un programma di analisi dei dati, era costretto a imparare ben cinque diverse sintassi: quella del sistema operativo, e quelle dei rispettivi programmi applicativi a cui era interessato. Da quando Windows si è imposto come il sistema operativo più diffuso sui *personal computer* tutti i programmi applicativi hanno cambiato impostazione e ora funzionano attraverso le cosiddette finestre di dialogo e i menù a tendina. Quindi gli utenti con il mouse aprono finestre di dialogo e scelgono quali comandi fare eseguire al programma. Indubbiamente la struttura a finestre di dialogo ha consentito l'accesso ai programmi informatici a un'estesa platea di utenti che prima era esclusa dalla necessità di dover imparare una o più sintassi per poter dialogare con il *computer*. Questo indiscutibile vantaggio presenta un rovescio della medaglia, o, se si preferisce, un prezzo da pagare: la maggior parte degli utenti sono del tutto inconsapevoli e incompetenti e, per così dire, più che usare un programma ne sono usati.

Ciò è vero soprattutto per i programmi applicativi, come Spss, che richiedono agli utenti delle competenze specialistiche al di là delle più o meno complesse competenze informatiche.

Infatti, seppure l'uso del programma non presenta particolari difficoltà, per una corretta utilizzazione dello stesso è comunque necessaria una preparazione di base sia statistica sia di metodologia e tecnica della ricerca. L'uso di un programma di analisi dei dati è metodologicamente corretto solo nella piena consapevolezza degli aspetti connessi al disegno della ricerca, nella comprensione degli output prodotti dalle elaborazioni statistiche e nel saperne valutare la coerenza e la pertinenza rispetto alle ipotesi iniziali e agli obiettivi cognitivi.

Proprio in virtù di tali necessità, l'eccessiva facilità d'uso che il sistema operativo Windows consente, si rivela ben presto un'arma a doppio taglio. L'uso inappropriato del programma si manifesta quando l'utente inconsapevole produce dei risultati che non è in grado di capire. A questo punto si tocca con mano l'inefficacia delle scorciatoie dovute all'eccessiva semplificazione dell'accesso ai comandi e alle procedure di analisi statistica e, di fatto, si realizza l'effetto di inutilità del lavoro svolto.

Per tali ragioni abbiamo sentito la necessità di dedicare il primo volume della nuova collana *La cassetta degli attrezzi. Strumenti per le scienze umane* proprio ad un manuale che illustri come si usa consapevolmente il più diffuso programma di analisi dei dati nelle scienze umane. Il patto che stipuliamo con i lettori è il seguente: se si accettano i costi di un percorso formativo che richiede uno sforzo maggiore si riceveranno dei benefici e delle gratificazioni in termini di gestione consapevole e autonoma del lavoro svolto sul programma di analisi dei dati.

Riprendendo la metafora espressa non sarete più usati dal programma ma sarete voi effettivamente ad usarlo.

Una volta apprese le regole essenziali della sintassi sarete in grado di impostare i comandi sulla base delle vostre specifiche esigenze e non sarete più costretti a ricorrere ai comandi impostati di default dal sistema che, dovendo soddisfare tutti gli utenti, il più delle volte non soddisfano realmente alcun utente.

Ci sono molti altri vantaggi nell'imparare a usare il programma attraverso la sintassi. Ad esempio, si avrà la possibilità di costruire una memoria del lavoro svolto e la possibilità di riuso dei programmi con le procedure di analisi dei dati in future sessioni di lavoro.

Il primo vantaggio è strategico quando il lavoro di analisi, come avviene di solito, è lungo e richiede diverse sessioni di lavoro in più giorni, o anche in più settimane. Può capitare facilmente di non ricordare cosa si è fatto all'inizio di un lavoro di analisi dei dati, o, comunque, ricostruire il lavoro svolto in diverse sessioni può essere molto faticoso. Se si è lavorato con la sintassi, si sarà costruita una memoria di tutte le operazioni effettuate e della loro sequenza temporale. Dando uno sguardo all'intero file che contiene la sintassi dei comandi eseguiti, sarà molto semplice ricostruire il percorso svolto fino ad un certo punto e capire da dove ripartire per sviluppare le ulteriori analisi. Inoltre, se come spesso capita, lavoriamo su diverse ricerche e quindi siamo nelle condizioni di dover svolgere le stesse operazioni di analisi su diverse matrici dei dati, si potranno riutilizzare i file sintassi costruiti nei precedenti lavori per condurre le stesse analisi sulle nuove matrici dei dati, una volta che si siano apportate le opportune modifiche agli stessi. In tal modo avremo il vantaggio di non dover cominciare da capo nel costruire il programma di analisi dei dati con un evidente risparmio di tempo.

Un ulteriore vantaggio nell'uso della sintassi consiste nella sua resistenza nel tempo. Come per tutti i programmi informatici, anche Spss è regolarmente aggiornato e sviluppato, tanto che annualmente la casa produttrice rilascia una nuova versione che presenta sempre qualche caratteristica diversa, e non sempre innovativa, rispetto alla precedente.

Sappiamo bene che l'innovazione dei prodotti informatici segue dei criteri economici e non sempre, a mio avviso raramente, l'ultima versione costituisce un effettivo miglioramento di quella precedente. Sta di fatto che i prodotti informatici devono costantemente essere aggiornati e così i poveri utenti ogni volta che acquistano la nuova versione devono dedicare parte del loro tempo per imparare le nuove funzioni e le nuove caratteristiche del programma.

Non tutti gli utenti hanno voglia di investire il tempo necessario per aggiornare lo stato delle loro conoscenze sull'uso del programma

e così il più delle volte succede che, pur avendo acquisito la nuova versione, continuano ad usare solo le caratteristiche che erano presenti nella versione precedente.

Conoscendo la sintassi del programma si rende di fatto ininfluenza la versione di Spss che attualmente è sul mercato. Si consideri che la sintassi è rimasta pressoché la stessa fin dalla prima versione del programma, quella che girava con il sistema operativo Dos.

Infine, ultimo ma non ultimo, l'uso della sintassi per alcune procedure di analisi consente la scelta di un numero maggiore di opzioni rispetto alla stessa procedura disponibile attraverso le finestre di dialogo. Pertanto, solo la sintassi consente una gestione flessibile e personalizzata delle procedure del programma e, quindi, un uso veramente professionale dello stesso.

A questo punto riteniamo di aver chiarito i motivi per i quali si è scelto di dedicare il presente volume alla programmazione di Spss attraverso la sintassi. Illustriamo quali conoscenze, competenze e strumenti si richiedono ai lettori per poter sfruttare pienamente i contenuti delle prossime pagine. Innanzitutto si dà per scontato che il lettore abbia una discreta familiarità con l'ambiente Windows e con i principali programmi applicativi di Microsoft Office come, ad esempio, Word ed Excel. Inoltre, si assume che il lettore sia in grado di aprire, copiare, salvare, spostare, nominare e rinominare files, creare cartelle, gestire le periferiche e soprattutto installare programmi sul proprio *personal computer*. Qualora mancassero queste competenze si rimanda ai manuali di Windows e ai manuali dei programmi Microsoft Office.

Rispetto alle competenze statistiche e metodologiche, si presume una preparazione assimilabile ad un corso universitario di primo livello in statistica e in metodologia della ricerca. Per ulteriori approfondimenti, nei paragrafi conclusivi di ciascun capitolo, saranno segnalati alcuni testi che si ritiene possano essere utili per applicare con cognizione di causa le procedure descritte nel testo.

Ovviamente si dà per scontato che il lettore disponga di un computer su cui sia installato il programma Spss.

Non importa quale versione sia installata, purché sia una versione compatibile con il sistema operativo Windows installato sul computer del lettore.

Come detto, la sintassi di Spss è sostanzialmente invariata nel corso del tempo, per cui non è necessario fare riferimento ad una versione particolare del programma per poter effettuare le procedure che saranno illustrate nei prossimi capitoli. Rispetto alle procedure di analisi e di pre-analisi dei dati che saranno oggetto del volume, quando non esplicitamente indicato, si sottintenderà che esse fanno parte del modulo base di una qualsiasi versione di Spss per Windows. Insieme al programma occorrerà avere a disposizione dei file dati che potranno essere creati dall'utente usando direttamente Spss o importati da altri programmi come, ad esempio, Excel. Questi file dati dovranno avere una struttura matriciale con in riga i casi e in colonna le variabili. Si tenga presente che il programma Spss mette a disposizione degli utenti diverse matrici dei dati (file con estensione .sav; vedi capitolo 2). Questi file vengono automaticamente copiati sul disco fisso degli utenti durante la procedura di installazione del programma. È quindi possibile aprirli e usarli per esercitarsi e per effettuare delle prove di analisi dei dati. Nel prossimo capitolo, dopo un'ulteriore illustrazione delle caratteristiche del programma, cominceremo proprio da come si apre o come si costruisce una matrice dei dati e come si effettuano su di essa le operazioni di analisi dei dati attraverso la costruzione di programmi di comandi usando la sintassi di Spss.



## 2. Il programma Spss

Spss è un sistema di programmi di calcolo scritti per analizzare dati relativi alle scienze umane. Esso presenta una notevole flessibilità per la lettura dei dati ed è corredato da una serie di procedure che permettono la loro trasformazione e manipolazione, e l'applicazione di un numero davvero imponente di procedure statistiche, dall'analisi monovariata fino alle più complesse analisi multivariate. Inoltre, è possibile produrre molti grafici semplici e complessi attraverso uno specifico modulo grafico.

Spss svolge queste funzioni usando tre tipi di comandi:

- 1) comandi operativi;
- 2) comandi di definizione, di manipolazione e di trasformazione dei dati;
- 3) comandi di procedure statistiche e/o grafiche.

I comandi operativi riguardano l'impostazione e la definizione del sistema piuttosto che i dati. Si usano per avere informazioni sul funzionamento di alcuni parametri del programma ed eventualmente per modificarli qualora non rispettino le esigenze degli utenti. Tutti questi comandi sono impostati su un valore di default scelto dai programmatori, ritenuto adatto alla maggior parte degli utenti. Naturalmente, qualora si ritenga necessario, è possibile cambiare l'impostazione di questi valori di default usando gli opportuni comandi operativi. Ad esempio, è possibile modificare il formato dell'output di Spss e scegliere la sua destinazione.

È anche possibile richiedere l'assistenza durante una sessione o chiudere una sessione di lavoro.

I comandi di definizione, manipolazione e trasformazione dei dati descrivono, etichettano e trasformano i dati, le variabili, i casi e l'intera matrice dei dati. Si usano per specificare a Spss dove e come leggere i dati e come etichettare le variabili e le loro modalità. I comandi di manipolazione dei dati rendono possibile la creazione di nuove variabili, la ricodifica dei valori delle variabili esistenti, l'identificazione dei valori da trattare come dati mancanti, etichettare le variabili e i valori ricodificati, normalizzare, standardizzare e deflazionare i valori delle variabili e qualsiasi altra possibile trasformazione matematica sugli stessi, selezionare (permanentemente o provvisoriamente) e/o ponderare i casi da usare in una particolare elaborazione.

I comandi di procedura fanno eseguire a Spss la produzione di rapporti, lo svolgimento di analisi statistiche e la memorizzazione dei dati nei file. Ci sono procedure che stampano e tracciano diagrammi dei dati, comparano gruppi e producono statistiche descrittive, bivariante e multivariate per qualsiasi tipo di variabili.

I comandi di Spss sono costituiti essenzialmente da: parole chiave (a loro volta distinguibili in parole chiave di comandi e parole chiave di sotto-comandi) e specifiche opzioni richieste dall'utente.

Il lavoro di Spss si svolge attraverso la creazione, esecuzione, manipolazione di file; di seguito si illustrano i principali tipi che è utile conoscere:

*File Dati:* contiene i dati che devono essere elaborati. Se è stato prodotto con Spss il suo formato è '*nome.sav*', dove '*nome*' rappresenta il nome del file scelto dall'utente; l'estensione '*.sav*' è automaticamente assegnata dal programma e identifica i file che contengono i dati e tutte le informazioni relative ad essi (ad esempio, le etichette delle variabili, le etichette delle modalità, e così via). L'insieme delle informazioni che riguardano le variabili presenti nella matrice dei dati costituisce il cosiddetto dizionario delle variabili. Ci sono molti modi per importare dei file dati prodotti da altri programmi. Nel paragrafo 2.3 presenteremo la procedura per importare file dati da Excel, in quanto riteniamo che la maggioranza dei lettori sia in grado di utilizzare il più familiare programma di Microsoft Office.

*File Sintassi:* è un file in cui sono memorizzati i comandi che l'utente intende eseguire per elaborare i dati, il suo formato è '*spss.sps*', dove '*spss*' rappresenta il nome del file e l'estensione '*.sps*' sono automaticamente as-

segnate dal programma. Per evitare che il file sintassi si cancelli alla chiusura del programma occorre salvarlo assegnandogli un nome diverso da quello impostato automaticamente dal programma.

*File Attivo:* è un file creato da Spss in cui vengono registrate tutte le informazioni relative alla struttura del file dati, alle variabili in esso contenute, alle variabili eventualmente ricodificate o calcolate nell'ambito della sessione in corso. È temporaneo, ossia viene cancellato automaticamente quando si chiude una sessione di lavoro.

*File di Output:* è un file creato da Spss in cui vengono presentati i risultati delle procedure applicate. Il file viene denominato '*spss.spo*', dove '*spss*' rappresenta il nome del file e l'estensione '*.spo*' sono automaticamente assegnate dal programma. Se non viene salvato con un nome diverso da quello assegnato dal programma, il file si cancella automaticamente alla chiusura di una sessione di lavoro.

*File di Sistema:* è un file in binario, copia del file attivo, registrato su disco, che può essere utilizzato anche in altre sessioni di lavoro. Il suo formato è '*spss.sys*', dove '*spss*' rappresenta il nome del file scelto e l'estensione '*.sys*' sono automaticamente assegnate dal programma.

Nei prossimi capitoli prenderemo in considerazione soprattutto i file dati, sintassi e output.

## **2.1 L'ambiente di lavoro Spss**

Il programma Spss si apre con le stesse modalità in uso per qualsiasi altro programma applicativo di Windows (vedi figura 2.1).

Nella figura 2.2 si presenta l'ambiente 'Editor dei dati' ('Data View' in inglese) di Spss una volta che il programma sia stato acceso dal menù programmi di Windows.

L'editor dei dati è costituito da una struttura a matrice nella quale sulle righe si dispongono i casi e sulle colonne le variabili. In alto a partire da sinistra si trovano dieci menù a tendina: i primi tre ('File', 'Modifica' e 'Visualizza') presentano le stesse funzioni di un qualsiasi programma applicativo di Microsoft Office. Ad esempio, dal menù 'File' si possono aprire i file, oppure si salvano assegnandogli