



DIPARTIMENTO DI SCIENZE POLITICHE E SOCIALI

Alessandro Marino

# Analisi dei dati mediante PSPP

Guida alla gestione di matrici *case* per *variabili* e all'uso delle tecniche statistiche *mono* e *bivariate* disponibili nel *free software* SPSS compatibile

*Dipartimento di Scienze Politiche e Sociali*  
*Alma Mater Studiorum Università di Bologna*

Bologna, Settembre 2015





Quest'opera è distribuita con Licenza  
Creative Commons Attribuzione - Non commerciale - Non opere derivate 3.0  
<http://creativecommons.org/licenses/by-nc-nd/3.0/deed.it>

Alessandro Marino

Analisi dei dati mediante PSPP: Guida alla gestione di matrici *casi* per *variabili* e all'uso delle tecniche statistiche *mono* e *bivariate* disponibili nel *free software* SPSS compatibile

Quest'opera è disponibile in AMS Acta, il deposito istituzionale per la ricerca dell'Alma Mater Studiorum  
Università di Bologna, alla pagina:  
<http://amsacta.unibo.it/4346>  
DOI 10.6092/unibo/amsacta/4346

# Contenuto

Questa guida è stata concepita come un manuale *tutor* rivolto direttamente agli studenti del Dipartimento di Scienze Politiche e Sociali della Scuola di Scienze Politiche dell'Università di Bologna. La platea può in ogni caso allargarsi agli universitari di altri Dipartimenti o Scuole impegnati in percorsi didattici o tesi di laurea richiedenti l'analisi statistica di dati quantitativi ( organizzati secondo la struttura della matrice dati rettangolare: casi per variabili ).

Il manuale [ nome: **AnaDataTUTOR.dpf** ] è scaricabile utilizzando il link del **Repository** indicato in uno qualsiasi dei siti Web dei docenti Roberto Cartocci, Michele Sapignoli, Paola Bordandini.

<http://www.unibo.it/SitoWebDocente/default.aspx?UPN=roberto.cartocci%40unibo.it>

<http://www.unibo.it/SitoWebDocente/default.aspx?UPN=michele.sapignoli%40unibo.it>

<http://www.unibo.it/SitoWebDocente/default.aspx?UPN=paola.bordandini%40unibo.it>

Può in ogni caso essere copiato su chiavetta USB nel laboratorio di Via Bersaglieri 4.

Il manuale *tutor* orienta l'utente nell'utilizzo di altri tre file, relativi al *software* e al materiale di ricerca scelto come supporto alla sperimentazione dell'analisi dei dati. I file sono:

- l'**eseguibile** necessario all'installazione del programma PSPP [ **pspp-084-20150427-32bits-Setup.exe** ];
- il pdf riprodotto il **questionario** strutturato impiegato nel sondaggio di riferimento [ **Qst\_08.dpf** ];
- il **file archivio** ( in PSPP ) contenente la matrice-dati e il registro delle variabili [ **Mat\_pol2008.sav** ].

Come acquisire il materiale per le esercitazioni.

**Eseguibile PSPP:** nel Laboratorio di Via Bersaglieri 4 o scaricandolo dalla rete internet.

**Questionario:** nel Laboratorio di Via Bersaglieri 4 o richiedendolo a [alessandro.marino@unibo.it](mailto:alessandro.marino@unibo.it) .

**File archivio in PSPP:** nel Laboratorio di Via Bersaglieri 4 o richiedendolo a [alessandro.marino@unibo.it](mailto:alessandro.marino@unibo.it) .

L'obiettivo del manuale *tutor* è quello di fornire la strumentazione minima in grado di consentire al neofita la coniugazione degli aspetti metodologico – statistici appresi in aula con la concreta pratica dell'analisi dei dati.

Il manuale *tutor* è diviso in tre sezioni:

**A** – Presentazione del programma PSPP: scarico e installazione su PC, apertura, chiusura; Menu di funzioni; esempio di creazione di un file-archivio.

**B** - Definizione operativa delle proprietà utilizzate nel sondaggio ( dalla mappa dei concetti alla costruzione della matrice dati );

**C** - Esercitazioni guidate di analisi monovariata e bivariata sui dati del sondaggio mediante PSPP .

# Indice

## Sez. A

A.1.	Presentazione di PSPP	pag. 1
A.2.	Modalità di installazione	pag. 1
A.3.	Apertura, modalità di visualizzazione e chiusura di PSPP	pag. 1
A.3.1.	Apertura	pag. 1
A.3.2.	Modalità di visualizzazione alternativa	pag. 2
A.3.3.	Chiusura di PSPP.	pag. 2
A.4.	Menu delle funzioni di PSPP	pag. 3
A.4.1.	<u>F</u> ile e <u>A</u> alyze.	pag. 3
A.4.2.	<u>D</u> ata e <u>T</u> ransform	pag. 4
A.5.	ESERCITAZIONE: esempio elementare di creazione di un archivio PSPP	pag. 5
A.5.1.	Dati inventati.	pag. 5
A.5.2.	Costruzione del registro delle variabili. Le variabili <i>genere</i> e <i>voto_met</i>	pag. 6
A.5.3.	Ultimazione del registro delle variabili e salvataggio del file PSPP	pag. 9
A.5.4.	Inserimento dei dati nella matrice	pag. 10
A.5.5.	Prova del Menu <u>A</u> alyze	pag. 10
A.5.6.	La finestra automatica di Output	pag. 11

## Sez. B

B.1.	Materiale utilizzabile per le esercitazioni	pag. 13
B.2.	Tema della ricerca, ambito e criteri di campionamento	pag. 13
B.3.	Mappa dei concetti	pag. 14
B.4.	Dalle proprietà alle variabili: le domande da inserire nel questionario	pag. 15

## Sez. C

C.1.	Uno sguardo preliminare sul file <i>Mat_pol2008.sav</i>	pag. 20
C.2.	ESERCIZIO n° 1 : analisi monovariata di alcune classificazioni	pag. 24
C.2.1.	Aggregazione delle categorie di una variabile: dalla D2 alla AggD2	pag. 26
C.2.2.	Aggregazione delle categorie mediante <i>Recode into Different Variables</i>	pag. 26
C.2.3.	Aggregazione delle categorie mediante la Sintassi di PSPP	pag. 28
C.3.	ESERCIZIO n° 2 : analisi monovariata di alcune scale	pag. 30
C.3.1.	Costruzione matematica di una nuova variabile: dalla D4 alla <i>Anni</i>	pag. 31
C.3.2.	Costruzione matematica di una nuova variabile mediante <u>C</u> ompute	pag. 32
C.3.3.	Costruzione matematica di una nuova variabile mediante la Sintassi di PSPP	pag. 33
C.4.	ESERCIZIO n° 3 : an. monovariata di una scala mediante <u>F</u> requencies e grafica	pag. 34
C.5.	Considerazioni propedeutiche agli ESERCIZI GUIDATI di analisi bivariata	pag. 37

C.5.1.	ESERCIZIO n° 4 : analisi bivariata tra due scale [ <i>Bivariate <u>C</u>orrelation</i> ]	<i>pag. 39</i>
C.5.2.	ESERCIZIO n° 5 : analisi bivariata tra più scale [ <i>Bivariate <u>C</u>orrelation</i> ]	<i>pag. 41</i>
C.5.2.1.	C.5.2.1. ESERCIZIO n° 6 : analisi bivariata per controllo ipotesi	<i>pag. 41</i>
C.5.3.	ESERCIZIO n° 7: an. bivariata tra scala e classificazione [ <i>Compare <u>M</u>eans</i> ]	<i>pag. 43</i>
C.5.4.	ESERCIZIO n° 8 : an. bivariata tra scala e classificazione [ <i>One Way <u>A</u>NOVA</i> ]	<i>pag. 45</i>
C.6.	Premesse allo studio congiunto di due classificazioni	<i>pag. 47</i>
C.6.1.	Significato dell'incrocio tra due classificazioni	<i>pag. 47</i>
C.7.	ESERCIZIO n° 9 : incrocio grezzo tra due classificazioni [ <i><u>C</u>rosstabs</i> ]	<i>pag. 48</i>
C.8.	ESERCIZIO n° 10. Ipotesi nulla: <i><u>C</u>rosstabs</i> con frequenze attese [ <i>Expected</i> ]	<i>pag. 50</i>
C.9.	ESERCIZIO n° 11. <i><u>C</u>rosstabs</i> con percentuali sul totale [ <i>Total</i> ]	<i>pag. 53</i>
C.10.	Percentuali di riga o colonna: strumenti efficaci nel controllo empirico di un modello asimmetrico di relazione tra due classificazioni	<i>pag. 54</i>
C.10.1.	ESERCIZIO n° 12: <i><u>C</u>rosstabs</i> con % di colonna [ <i>Column</i> ]	<i>pag. 55</i>

## SEZIONE A

### A.1. Presentazione di PSPP.

PSPP è un programma creato dalla Free Software Foundation all'interno del progetto di sviluppo del sistema operativo GNU, tuttora in evoluzione [ per info: <https://www.gnu.org/software/pspp/> ]. Esso consente l'esecuzione di numerose procedure statistiche secondo le modalità operative di SPSS ( Statistical Package for the Social Sciences ), capostipite dei programmi per l'analisi dei dati .

PSPP assicura il massimo livello di efficienza su piattaforma Linux. Le versioni a 32 e 64bits adattate ai sistemi operativi Windows presentano invece una qualità grafica inferiore e mancano parzialmente o totalmente di alcune *utilities* proprie di SPSS. Nonostante tali lacunosità, allo scopo di assicurare la massima diffusione fra i potenziali utenti, viene reso disponibile il programma PSPP a 32bits, funzionante con sistemi operativi Windows XP, Vista e Seven .

### A.2. Modalità di installazione.

1-copiare/salvare sul desktop del computer il file eseguibile [ *pspp-084-20150427-32bits-Setup.exe* ] ;

2-cliccare due volte sul nome del file eseguibile. **Esiti** :

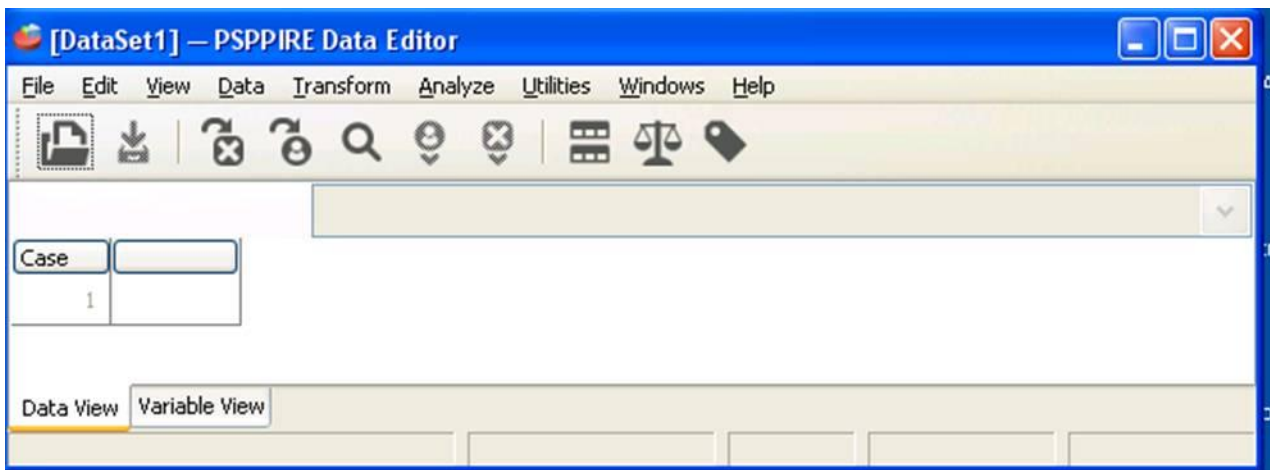
- automatica installazione di PSPP nella directory C:\Programmi;
- creazione sul desktop dell'icona di collegamento →



### A.3. Apertura, modalità di visualizzazione e chiusura di PSPP.

#### A.3.1. Apertura.

Cliccare due volte sull'icona di collegamento. **Esito** :



#### Osservazioni

La finestra di PSPP si apre nella modalità di **visualizzazione dati** ( *Data View* , con sottolineatura **gialla** in basso a sinistra ). Nell'esempio, la mancata apertura di un file PSPP pre-esistente determina l'assenza di informazioni.

### A.3.2. Modalità di visualizzazione alternativa.

Cliccare sulla voce Variable View (alla destra di Data View) . **Esito:**



#### Osservazioni

La finestra si apre nella modalità **visualizzazione variabili** . E' il registro di PSPP: la parte del programma in cui sono stabiliti i criteri necessari all'identificazione e gestione delle variabili e dei relativi dati inseriti nella matrice. In assenza dell'apertura di un archivio dati pre-esistente, tutti i campi del registro sono vuoti. In ogni caso, al di sopra di ogni campo vuoto sono esplicitate le caratteristiche a cui ogni variabile sarà associata. Di seguito sono elencate succintamente le voci principali, ciò che esse significano e la loro condizione di necessità.

**Name:** contiene il *nome* assegnato dal ricercatore ad ogni singola variabile [ **necessario** ];

**Type:** indica la natura *numerica* o *alfanumerica* dei dati riferibili alla variabile [ **necessario** ];

**Width:** determina il numero di spazi occupati da ogni possibile dato della variabile [ **necessario** ];

**Decimals:** il numero di cifre *decimali* evidenziato [ **necessario**. PSPP assegna automaticamente due decimali, modificabili opzionalmente dall'utente ] ;

**Label:** *etichetta lessicale* associabile al nome della variabile [ **opzionale** ];

**Value Labels:** *etichette lessicali* riferibili a ciascuna *modalità* di una variabile qualitativa [ **opzionale** ];

**Missing Values:** *codici numerici* utilizzati per indicare l'assenza della risposta [ **opzionale** ].

Nel paragrafo **A.5.** verrà fatto un esempio elementare del funzionamento del registro in fase di creazione di un archivio PSPP, mentre nella sezione **C** si vedrà in dettaglio come il registro funziona relativamente alla base dati [ *Mat\_pol2008.sav* ] utilizzata per le esercitazioni.

### A.3.3. Chiusura di PSPP.

Cliccare sulla voce **File** in alto a sinistra →

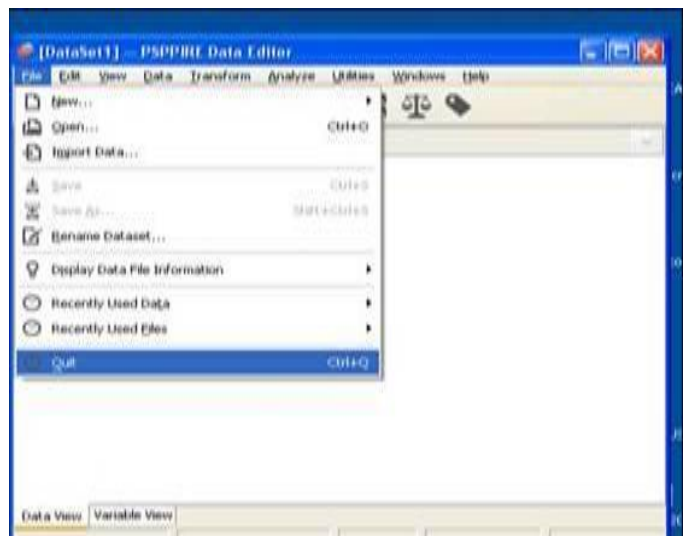
**Esito:**

apertura della sotto-finestra →

nella quale evidenziare e cliccare l'ultima voce,

**Quit** →

Il programma si chiude.



## A.4. Menu delle funzioni di PSPP.

Il Menu delle funzioni è costituito da nove voci elencate nella prima riga al di sotto della fascia blu di ogni tipo di finestra PSPP. Ogni funzione è richiamabile con un clic del mouse.



Dopo il clic si apre una sotto-finestra in cui sono elencate opzioni attivabili direttamente o indicanti a loro volta ( → ) due o più opzioni fra cui scegliere. Tutte le opzioni sono anch'esse attivabili con un clic.

Le funzioni indispensabili sono **File** e **Analyze** . Altre due funzioni importanti, utilizzabili opzionalmente sulla base di esigenze specifiche, sono **Data** e **Transform**.

### A.4.1. **File** e **Analyze**.

**File** . Le opzioni più usate di questa funzione sono:

#### **New** .

Consente di aprire un file di sintassi ( in cui scrivere i comandi in alternativa al Menu grafico ) o di aprire un nuovo file in parallelo a quello già aperto.

#### **Open** .

Apre un file, dopo avere:

- indicato dove esso è memorizzato sul PC ed
- evidenziato il suo nome .

#### **Save** .

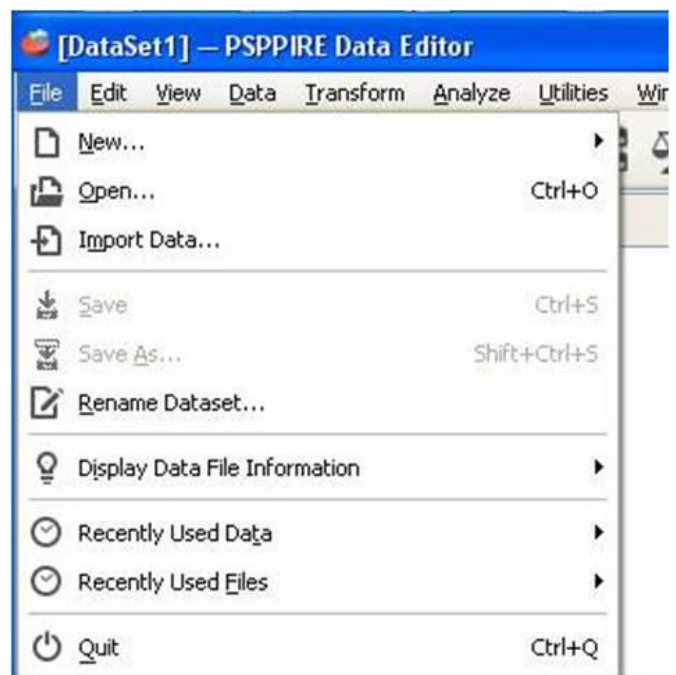
Salva il file su cui si sono operate delle modifiche, mantenendo lo stesso nome e lo stesso percorso di memorizzazione sul PC.

#### **Save As** .

Salva il file su cui si sono operati interventi nella matrice-dati e/o sulle variabili, modificando il nome e/o il percorso di memorizzazione sul PC.

#### **Quit** .

Provoca la fuoruscita dal programma.





## Analyze .

Rappresenta il portone d'ingresso per richiedere qualsiasi procedura statistica. Le opzioni offerte da questa funzione, quando sono affiancate dalla freccetta (→) danno la possibilità di scegliere fra modalità alternative dello stesso tipo di procedura. Le opzioni più utilizzate sono:

**Descriptive Statistics** (→). Esse comprendono:

- **Frequencies** . Calcolo delle distribuzioni di frequenza, con l'aggiunta di Media e Scarto Tipo;
- **Descriptives** . Calcolo dei valori medi e di variabilità per variabili quantitative ( scale );
- **Crosstabs** . Tavole di contingenza tra variabili qualitative.

**Compare Means** (→).

Analisi della varianza esterna relativa alla relazione tra una scala e una classificazione, partendo dal calcolo della media della scala su ciascun gruppo di casi riconducibile alle singole categorie della classificazione.

E' possibile ottenere la procedura ANOVA con produzione di  $Eta^2$ , così come varie forme di T test per piccoli campioni.

**Bivariate Correlation** .

Correlazione fra variabili quantitative ( scale ).

**Regression** (→).

Analisi della regressione lineare o della binaria logistica.

**Factor Analysis** .

Analisi delle componenti principali.

### A.4.2. Data e Transform .

**Data** . Le opzioni più utilizzate di questa funzione sono:

**Sort Cases** .

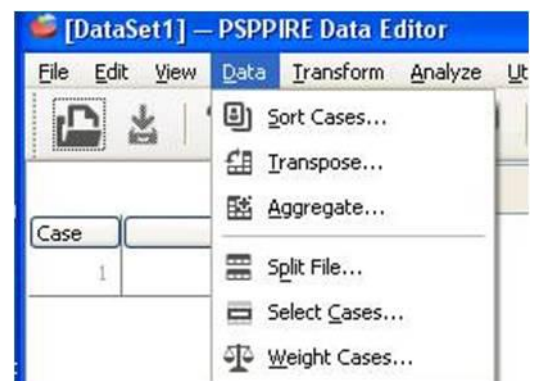
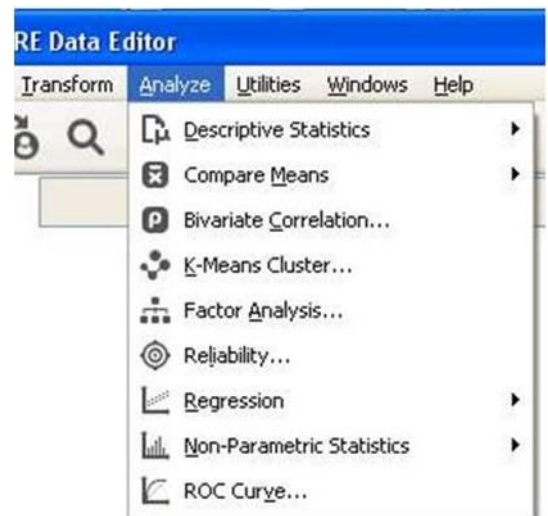
Riordina i casi secondo i valori crescenti o decrescenti di una o più variabili.

**Select Cases** .

Seleziona un insieme di casi rispetto al totale sulla base del soddisftimento di una condizione logica coinvolgente una o più variabili ( ad. Es. : maschi con più di 35 anni ).

**Aggregate** .

Aggiunge nuove variabili ad un file attivo.



**Transform**. Le opzioni più utilizzate di questa funzione sono:

**Compute** .

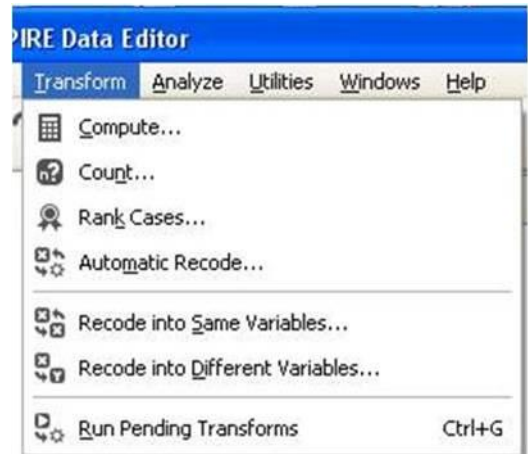
Crea o modifica una variabile in base ad un'operazione matematica coinvolgente una o più variabili.

**Recode into Same Variables** .

Modifica una variabile aggregando due o più categorie della stessa, riducendo così il numero di modalità.

**Recode into Different Variables** .

Stesse modifiche, però riprodotte in una nuova variabile.



## A.5. ESERCITAZIONE: esempio elementare di creazione di un archivio PSPP.

In generale, la creazione ex novo di un file archivio PSPP presuppone l'esistenza di un volume di informazioni attribuibili ad un certo numero di referenti empirici ( casi di una ricerca ), su ciascuno dei quali siano state registrate in modo sistematico ( su un qualsiasi supporto cartaceo o informatico ) gli stati corrispondenti ad una serie di proprietà.

Dal punto di vista operativo, occorre eseguire due fasi cronologicamente distinte:

a) la costruzione del registro delle variabili [ vedi **A.5.2.** ];

b) l'inserimento dei dati nella matrice [ vedi **A.5.4.** ] .

### A.5.1. Dati inventati.

Si supponga di porre alcune domande a nove studenti, presi singolarmente al termine dell'esame orale del corso triennale di Metodologia. Prima di iniziare l'intervista, viene deciso di registrare l'informazione sul sesso, scrivendo sull'apposito modulo - alla voce *genere* - il codice 1 o 2 se in presenza rispettivamente di una studentessa o di uno studente. Ultimata l'operazione, inizia l'intervista.

Qui sotto, la sostanza delle domande e i criteri di codifica stabiliti per registrare le risposte:

- voto in trentesimi conseguito → da 18 a 30; con 32 in luogo del 30 e lode e 0 per l'insufficienza;
- tipo di esaminatore → 1 - titolare del corso; 2 – altro membro della commissione;
- soddisfazione esito esame → 0 – insoddisfatto; 1 – poco s. ; 2 – abbastanza s. ; 3 – molto soddisfatto;
- condizione rispetto agli studi → 1 – in corso; 2 – fuori corso;
- logistica abitativa → 1 – bolognese; 2 – fuori sede domiciliato a Bologna; 3 – fuori sede pendolare
- n° esami sostenuti →  $n$  ;
- media esami in classi → 1 – da 18 fino a 23,5 ; 2 – da 23,6 a 26,5 ; 3 – sopra 26,5 .

Di seguito le ipotetiche risposte, codificate sul modulo cartaceo secondo i criteri sopra indicati:

MODULO CARTACEO UTILIZZATO DALL'INTERVISTATORE

genere	voto_met	esaminatore	soddisfazione	stato_corso	abita	n_esami	media
1	24	2	2	1	1	5	2
2	19	1	1	2	2	4	1
2	28	1	3	1	2	6	3
1	30	2	3	1	1	5	3
2	24	2	2	2	1	6	2
1	23	2	1	1	3	4	2
1	32	1	3	1	1	5	3
1	27	2	2	1	2	5	3
2	0	1	0	2	1	4	1

La griglia raffigura una matrice-dati rettangolare.

La riga sovrastante i dati riporta i **Nomi** delle proprietà. Gli stessi nomi saranno poi usati per identificare le variabili nel file PSPP.

Il contenuto della matrice-dati, esplorabile orizzontalmente ( **per riga** ) e verticalmente ( **per colonna** ), indica che sono osservabili:

- **per riga**, i dati relativi agli stati delle otto proprietà registrate su uno specifico caso;
- **per colonna**, i dati relativi agli stati rilevati su tutti i casi per la medesima proprietà.

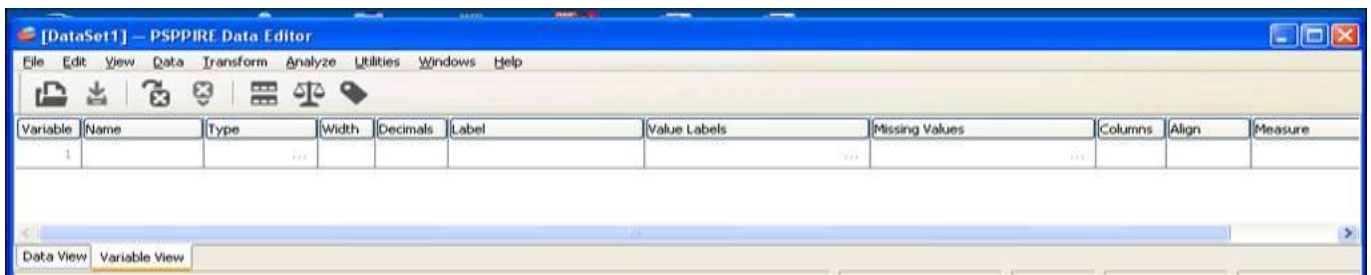
### A.5.2. Costruzione del registro delle variabili. Le variabili *genere* e *voto\_met* .

Si eseguono in successione i passi sottostanti:

- Aprire PSPP, cliccando sull'icona di collegamento →



- Entrare in modalità visualizzazione variabili cliccando sul bottone **Variable View**

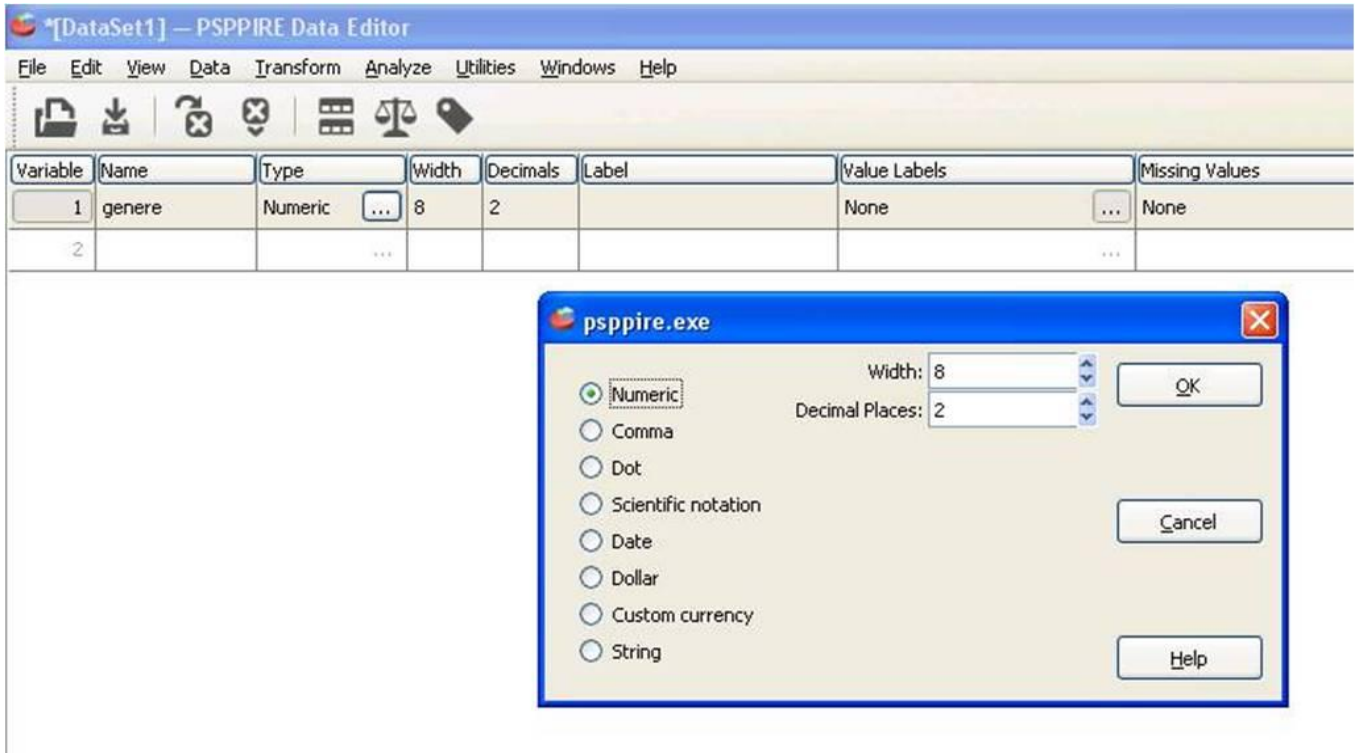


- Cliccare una volta nel campo vuoto **Name** ( alla destra del campo automatico *Variable* ).

**Esito:** nel campo compare il termine Var001 : è il nome assegnato in automatico da PSPP alla prima variabile, in assenza di un nome attribuito dall'utente.

- Sostituire il termine Var001 sovra-scrivendo il nome deciso per identificare la prima variabile: **genere**
- Passare col mouse sul campo **Type** , cliccare sui tre puntini ...

**Esito:** apertura della sotto-finestra di seguito illustrata



La sotto-finestra propone in automatico il tipo *numerico*, associando ad esso otto spazi e due decimali ( sotto-campi **Width** e **Decimals Places** ). Essendo la variabile genere registrabile con i soli codici 1 e 2 [ bastevoli di un solo spazio/colonna e nessun decimale ] sarà pratico ridurre a **1** e **0** i valori dei sotto-campi, cliccando più volte sulla freccetta ( ↓ ) fino ad ottenere le cifre desiderate. Ultimata l'operazione, cliccare sul bottone **OK** . PSPP registrerà così la prima variabile, identificandola tramite il nome *genere* , con valori possibili di tipo *numerico* compresi nella gamma 0→9 e privi di decimali .

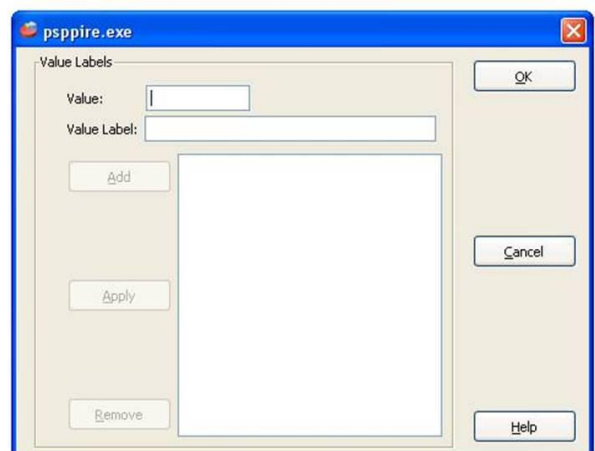
- Passare al campo **Label** e inserire l'etichetta **Sesso dell'esaminato**

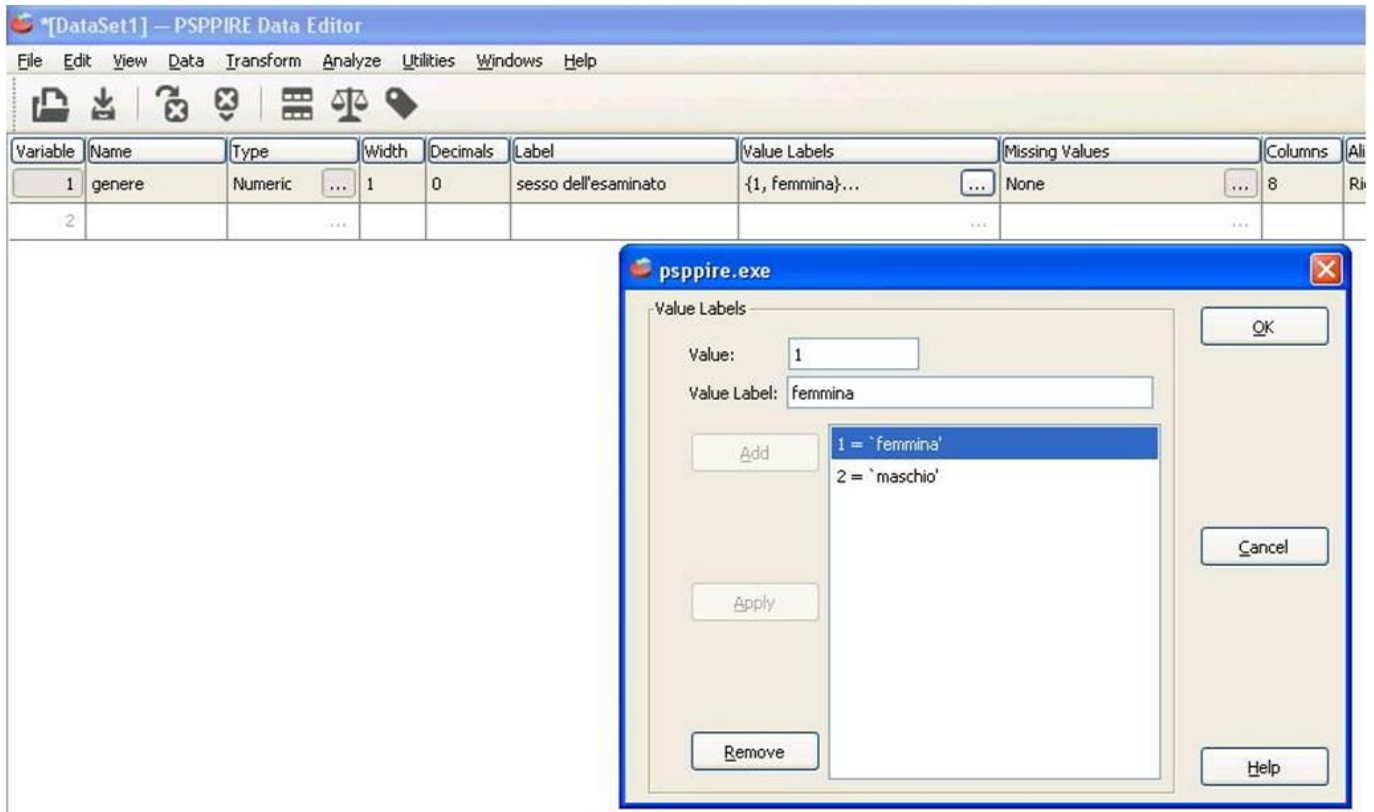
- Passare al campo **Value Labels** e cliccare una volta sui tre puntini a destra ... .

**Esito:** apertura della sottofinestra in cui inserire le etichette relative a ciascun codice previsto .

Procedere nel seguente modo:

- inserire il valore **1** nella location **Value** quindi
- passare nella location **Value Label** e scrivere *femmina*
- cliccare quindi sul bottone **Add** ( ottenendo in tal modo l'inserimento nel riquadro più grande di 1 = 'femmina' )
- tornare su **Value** e inserire il valore **2** quindi
- scrivere *maschio* nella location **Value Label** e
- cliccare nuovamente sul bottone **Add** .



**Esito:**

- A questo punto è sufficiente cliccare sul bottone **OK** . La sotto-finestra si chiude, le etichette sono state inserite nel registro e si è quindi pronti per definire le caratteristiche della seconda variabile: il voto conseguito all'esame di Metodologia.

Si procederà analogamente a quanto visto per la variabile *genere*. Dunque:

- nel campo **Name** verrà scritto il nome stabilito: *voto\_met* ;
- nel campo **Label** verrà scritta l'etichetta: *Voto di Metodologia*;
- il campo **Value Label** sarà invece lasciato vuoto, trattandosi di una scala numerica di cui non avrebbe senso etichettare le possibili modalità da 18 a 30, compreso il 32 a rappresentare il 30 e lode .

Si dovrà tuttavia tenere conto di aver deciso di codificare con **0** l'eventuale valutazione insufficiente. In questo senso sarà utile indicare al programma che tale valore numerico, se presente nella matrice dei dati, andrà considerato opportunamente. Da una parte sarà infatti utile conoscere quanti studenti sul totale non hanno superato l'esame, dall'altra sarà necessario escludere il valore 0 degli studenti respinti qualora si intenda calcolare la media dei voti di Metodologia. PSPP consente questo duplice trattamento attraverso l'attribuzione, nel registro, di un determinato codice ( nell'esempio, 0 ) definibile dall'utente come *dato mancante* riferito ad una specifica variabile ( nell'esempio, *voto\_met* ) .

La funzione è denominata **Missing Values** . Procedere come segue:

- sulla riga del registro relativa alla variabile *voto\_met*, posizionarsi nella cella corrispondente alla colonna **Missing Values** e

- cliccare sui tre puntini a destra ...

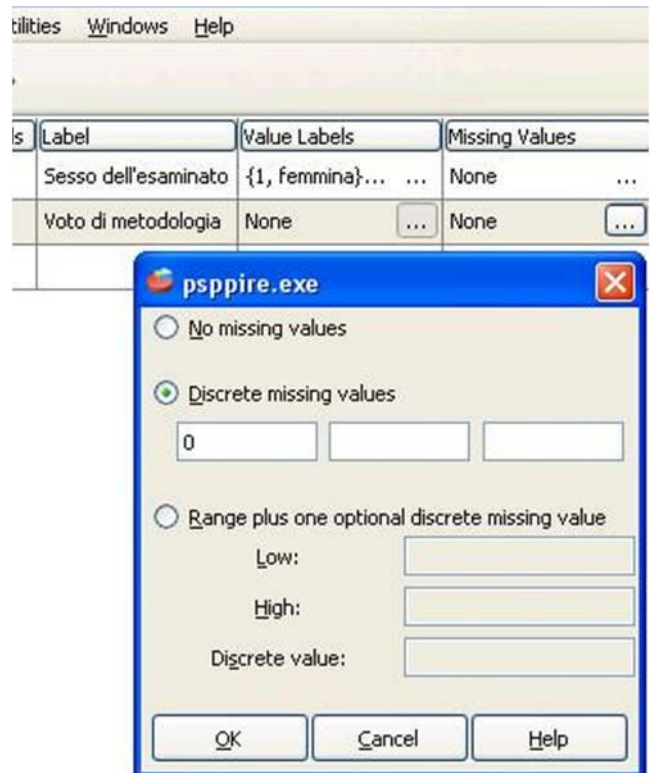
**Esito:** si apre una sottofinestra in cui occorre:

- flaggare la voce **Discrete missing values**, quindi:

- inserire il valore numerico 0 nel primo dei tre campi vuoti disponibili e infine:

- cliccare sul bottone **OK**.

A questo punto, la variabile *voto\_met* avrà il valore numerico 0 stimabile sia come codice "buono" sia come codice determinante l'esclusione del caso in occasione di calcoli statistici o incroci di *voto\_met* con altre variabili.



### A.5.3. Ultimazione del registro delle variabili e salvataggio del file PSPP.

Le rimanenti sei variabili da specificare nel registro non presentano valori da definire come mancanti. Oltre a ciò, a parte la proprietà enumerabile *n\_esami*, le altre variabili *esaminatore*, *soddisfazione*, *stato\_corso*, *abita* e *media* potranno essere etichettate nelle specifiche modalità previste.

Al di là della soluzione qui proposta, ogni utente potrà sbizzarrirsi nella definizione delle etichette.

Una volta completata la costruzione del registro delle variabili, **sarà opportuno procedere al salvataggio** di quanto fatto. L'apertura generica di PSPP determina infatti l'esistenza di un file temporaneo che, se si chiude la sessione senza salvarlo, viene perduto. Occorrerà pertanto:

- aprire il Menu **File** e

- selezionare l'opzione **Save As**

decidendo quale nome assegnare al file e in quale area/cartella del PC memorizzarlo.

Ultimata l'operazione, il programma continuerà a mostrare la finestra attiva, indicando però il nome del file scelto, posizionato a sinistra della fascia superiore blu [ *provafile.sav* ]

Variable	Name	Type	Width	Decimals	Label	Value Labels	Missing Values	Columns	Align	Measure
1	genere	Numeric ...	1	0	Sesso dell'esaminato	{1, femmina}...	None ...	8	Right	Scale
2	voto_met	Numeric ...	2	0	Voto di metodologia	None ...	0 ...	8	Right	Scale
3	esaminatore	Numeric ...	1	0	Qualifica dell'esaminatore	{1, Titolare Corso}...	None ...	8	Right	Scale
4	soddisfazione	Numeric ...	1	0	Grado di soddisfazione per esito esame	{0, insoddisfatto}...	None ...	8	Right	Scale
5	stato_corso	Numeric ...	1	0	condizione rispetto agli studi	{1, in corso}...	None ...	8	Right	Scale
6	abita	Numeric ...	1	0	logistica residenza&università	{1, risiede a bologna}	None ...	8	Right	Scale
7	n_esami	Numeric ...	2	0	numero esami sostenuti	None ...	None ...	8	Right	Scale
8	media	Numeric ...	1	0	valore della media in classi	{1, 18 --> 23,5}...	None ...	8	Right	Scale

Il completamento del file necessita a questo punto dell'inserimento delle codifiche delle risposte, scritte a mano sul modulo cartaceo utilizzato dall'intervistatore e illustrato nel paragrafo **A.5.1.** .

#### A.5.4. Inserimento dei dati nella matrice.

L'inserimento dei codici relativi alle risposte alle singole domande è semplice. Una volta aperta la finestra PSPP in modalità *visualizzazione dati*, è sufficiente posizionarsi con il mouse nella cella della prima riga sottostante la variabile *genere* ed inserire il codice 1 relativo al primo intervistato. Utilizzando comodamente la freccetta (→) della tastiera ci si sposterà nel campo della variabile successiva, *voto\_met*, e si inserirà il numero 24. Completato il primo caso, si passerà alla riga sottostante, ripetendo gli inserimenti delle risposte per il secondo intervistato.

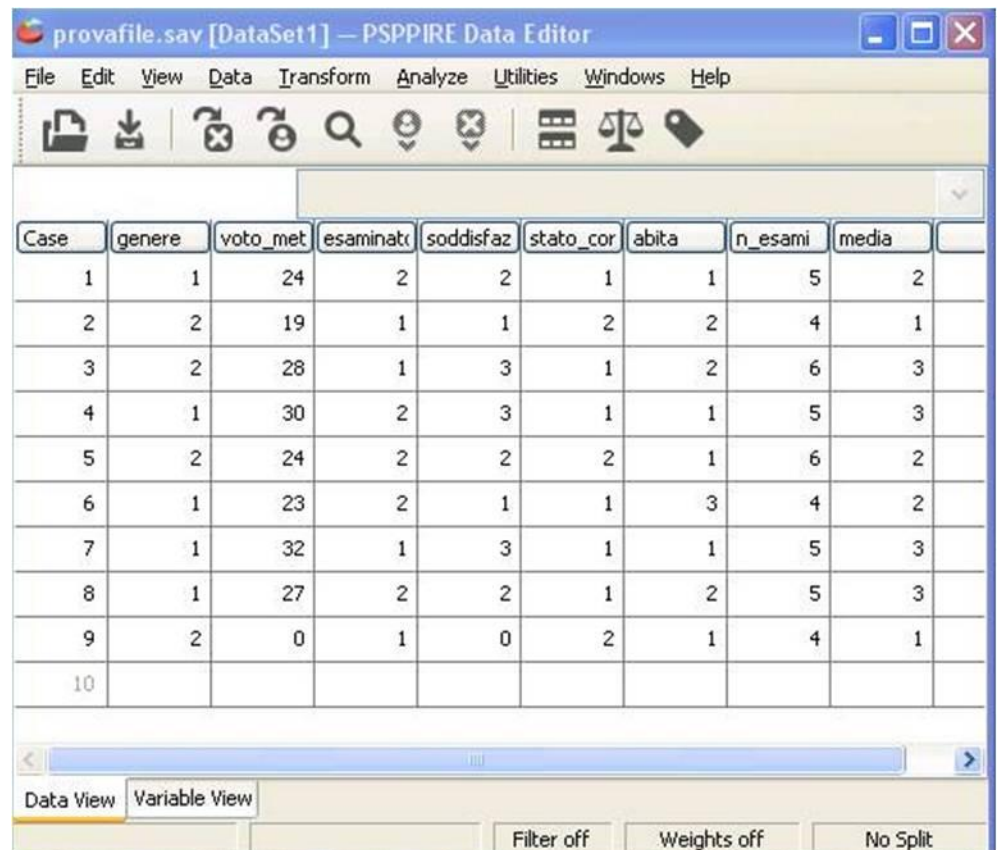
Reiterando il meccanismo per i rimanenti casi si completerà la costruzione della matrice-dati, come a lato evidenziato ...

→

Notare la voce **Case**, prima colonna della matrice.

E' la variabile automatica con cui PSPP assegna un numero identificativo ad ogni riga inserita nella matrice, come Excel con i suoi fogli di lavoro .

**Case** non fa quindi parte del lotto di variabili riferite ai casi e non può essere utilizzata né da sola né in relazione ad altre variabili.



The screenshot shows the PSPP Data Editor window titled 'provafile.sav [DataSet1] - PSPP Data Editor'. The window contains a data matrix with the following columns: Case, genere, voto\_met, esaminati, soddisfaz, stato\_cor, abita, n\_esami, and media. The data is as follows:

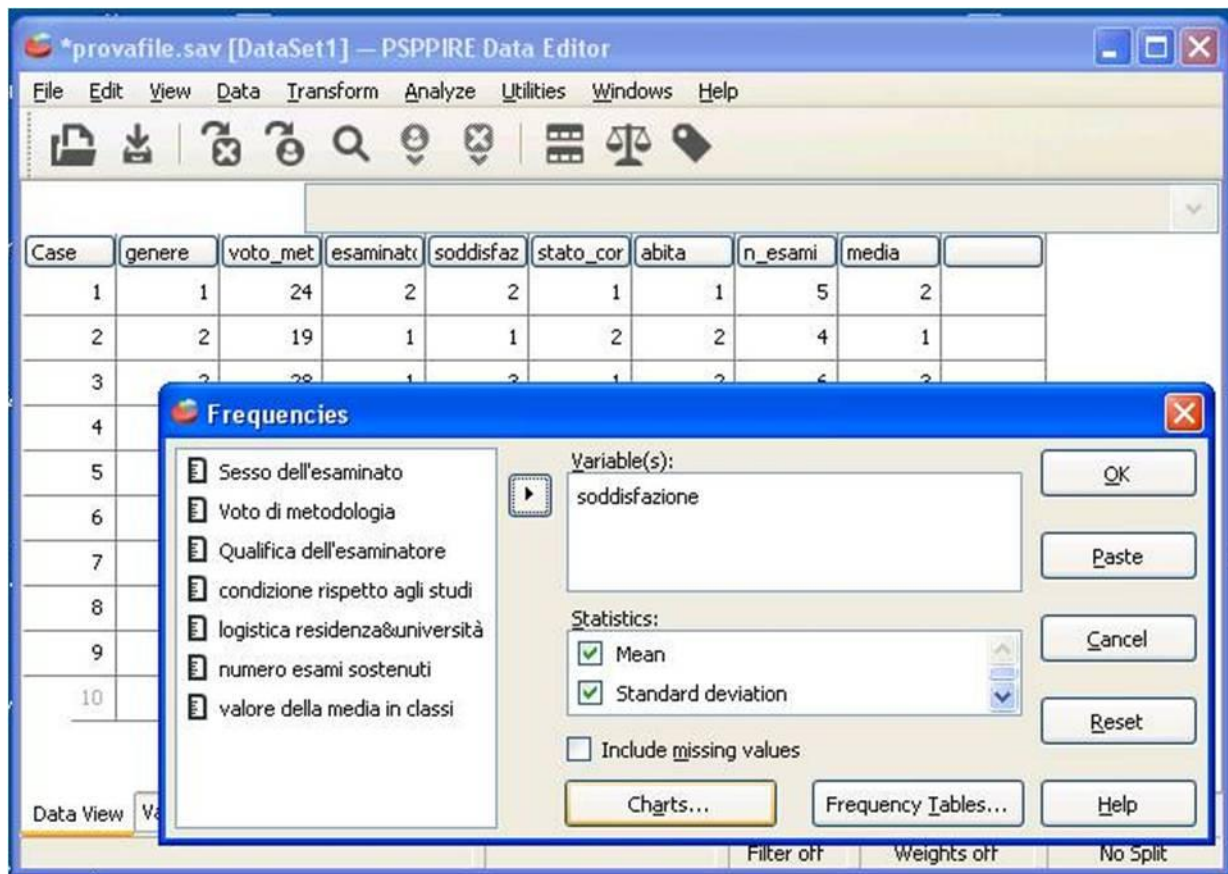
Case	genere	voto_met	esaminati	soddisfaz	stato_cor	abita	n_esami	media
1	1	24	2	2	1	1	5	2
2	2	19	1	1	2	2	4	1
3	2	28	1	3	1	2	6	3
4	1	30	2	3	1	1	5	3
5	2	24	2	2	2	1	6	2
6	1	23	2	1	1	3	4	2
7	1	32	1	3	1	1	5	3
8	1	27	2	2	1	2	5	3
9	2	0	1	0	2	1	4	1
10								

The window also shows a menu bar (File, Edit, View, Data, Transform, Analyze, Utilities, Windows, Help) and a toolbar with various icons. At the bottom, there are tabs for 'Data View' and 'Variable View', and buttons for 'Filter off', 'Weights off', and 'No Split'.

#### A.5.5. Prova del Menu Analyze .

Il file di prova è pronto. Per semplicità, si chiederà la distribuzione di frequenza e il calcolo della Media e dello Scarto Tipo della variabile *soddisfazione*. Essa è infatti da ritenersi una pseudo – scala poiché, come nella Likert, le modalità di risposta - costituite da picchetti semantico linguistici - assicurano il requisito dell'ordinalità, intrinseco ai valori di una scala, ma approssimano solamente quello dell'equidistanza fra i picchetti. Ha quindi senso esplorare questo tipo di variabile sia qualitativamente che, con la dovuta cautela, quantitativamente. Si proceda pertanto come segue:

- selezionare **Analyze** (→) **Descriptive Statistics** (→) **Frequencies** ( Si apre la relativa sotto-finestra );
- scegliere - dalla lista a sinistra – la variabile desiderata ( espressa con la sua etichetta: *Grado di soddisfazione per l'esito dell'esame* ) e spostarla nella *location* di destra con un clic sulla freccetta [ → ].



**NB.** La procedura **Frequencies** propone automaticamente, oltre alla distribuzione di frequenza, anche il calcolo della Media e dello Scarto tipo ( Standard Deviation ). Ovviamente, tale automatismo è utile in presenza di scale o pseudo – scale, ma diventa privo di senso quando la variabile è una classificazione in senso stretto ( come ad es. la variabile *esaminatore* ). In tali casi occorrerà preventivamente deflaggare con un clic i *Check Mark* [ ✓ ] corrispondenti.

Tornando all’esempio della variabile *soddisfazione* , sarà sufficiente cliccare sul bottone **OK** per attivare l’esecuzione della procedura. Il risultato è illustrato nel paragrafo successivo.

#### A.5.6. La finestra automatica di Output.

PSPP ha eseguito la procedura statistica richiesta. In concreto, si è posizionato sulla colonna corrispondente alla variabile *soddisfazione* e ha contato, riga per riga, quante volte ricorrono nelle nove celle i codici numerici 0, 1, 2 e 3 . L’esito di questi conteggi viene quindi organizzato in una tabella riepilogativa, accompagnata da una seconda in cui sono indicati i valori della Media e dello Scarto Tipo. Tutte queste informazioni vengono scaricate nella sottostante finestra di Output .



The screenshot shows the PSPPIRE Data Editor interface. The main window displays the 'Output - PSPPIRE Output Viewer' window, which contains the following information:

**Command Syntax:**

```

FREQUENCIES
FREQUENCIES
/VARIABLES= soddisfazione
/FORMAT=AVALUE TABLE.

```

**Grado di soddisfazione per esito esame**

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
insoddisfatto	0	1	11,11	11,11	11,11
poco soddisfatto	1	2	22,22	22,22	33,33
abbastanza soddisfatto	2	3	33,33	33,33	66,67
molto soddisfatto	3	3	33,33	33,33	100,00
<i>Total</i>		9	100,0	100,0	

**Grado di soddisfazione per esito esame**

<i>N</i>	<i>Valid</i>	9
	<i>Missing</i>	0
<i>Mean</i>		1,89
<i>Std Dev</i>		1,05
<i>Minimum</i>		,00
<i>Maximum</i>		3,00

Prima di commentare il risultato, è utile ricordare che la finestra di Output rimane attiva fino al termine della sessione di PSPP, a meno che l'utente decida di chiuderla preventivamente con un clic sulla **X** posizionata nell'angolo in alto a destra. Oltre a ciò, attraverso la voce **File**, il contenuto della finestra di Output può essere stampato oppure esportato in un file di testo [ formato **\*.txt** ] apribile da Word.

**Commento alla distribuzione di frequenza.** L'Output riporta, come prima informazione, la sintassi utilizzata dal programma per eseguire quanto richiesto attraverso il Menu grafico. Al di sotto della sintassi dei comandi compaiono le due tabelline riepilogative. Ognuna reca in testa l'etichetta attribuita dall'utente al nome della variabile.

La tabellina relativa alla distribuzione di frequenza visualizza le etichette (*Value Label*) attribuite alle modalità di risposta, disposte secondo l'ordine crescente dei codici (*Value*) stabiliti per la variabile *soddisfazione*. Segue il conteggio (*Frequency*) dei casi per ciascun valore possibile (nell'esempio si contano: uno studente insoddisfatto, due poco soddisfatti, e tre abbastanza o molto soddisfatti). Il totale fatto a 100 produce nella colonnina a fianco (*Percent*) la percentualizzazione di questi conteggi. Tale percentuale viene ricalcolata (*Valid Percent*) escludendo eventuali casi mancanti. Nell'ultima colonna (*Cum Percent*) compaiono infine le percentuali cumulate, utili da osservare ogniqualvolta le modalità della variabile sono ordinate. Nell'esempio, si potrà dire che meno del 34% degli studenti intervistati si ritiene poco o per niente soddisfatto dell'esito dell'esame.

**Commento al calcolo della Media e dello Scarto Tipo.** La tabellina riporta, nell'ordine: il numero di casi validi; il numero di eventuali casi mancanti; il valore della Media; quello dello Scarto Tipo; i valori minimi e massimi (*range* della variabile *soddisfazione*).

Questo paragrafo chiude l'esercitazione sul file di prova. Lo studente può comunque "testare" il programma, operando altre richieste prima di passare al materiale di ricerca vero e proprio.

## SEZIONE B

### B.1. Materiale utilizzabile per le esercitazioni.

Le esercitazioni mediante PSPP illustrate nel manuale prevedono il seguente materiale di riferimento [ scaricabile a lezione o richiedendolo via mail all'autore ]:

- la matrice dati [ *Mat\_pol2008.sav* ];
- la copia del questionario strutturato [ *Qst\_08.dpf* ] impiegato nella fase di intervista.

La matrice dati e il questionario sono relativi ad un mini sondaggio riguardante le opinioni politiche e il comportamento di voto dell'elettorato nelle settimane antecedenti le Elezioni del 13-14 Aprile 2008 per il rinnovo delle rappresentanze parlamentari di Camera e Senato.

Il sondaggio venne proposto nell'ambito delle attività formative per l'acquisizione dei crediti di Laboratorio II, e coinvolse 24 studenti frequentanti il biennio specialistico di Scienza Dell'Organizzazione e del Governo. Gli studenti effettuarono complessivamente 271 interviste, costruirono la matrice-dati relativa alle risposte degli intervistati alle singole domande e procedettero all'analisi statistica dei dati.

### B.2. Tema della ricerca, ambito e criteri di campionamento.

#### Il fallimento del Prodi II.

La situazione politica nel 2007 e nei primi mesi del 2008 fu caratterizzata dal progressivo sfaldamento degli equilibri interni al secondo governo guidato da Romano Prodi, insediatosi nel Maggio 2006 . Il Prodi II vedeva, oltre al PD (sorto il 14 Ottobre 2007 dalla confluenza di DS e Margherita), la partecipazione di Rifondazione Comunista e dei Radicali italiani, dell'UDEUR, di Italia dei Valori e di altre formazioni liberal democratiche. Nonostante fosse ancora nitido il ricordo del fallimento del Governo Prodi I, affossato il 9 Novembre del 1999 dal voto di sfiducia di Fausto Bertinotti, con autolesionistico tempismo il Centro Sinistra [ d'ora in avanti CSX ] implose per la seconda volta. Il sottile equilibrio fra le diverse componenti venne infatti rotto il 24 gennaio 2008 dal voto di sfiducia al Senato ad opera di Clemente Mastella (ex ministro della Giustizia dimissionario, su cui gravavano otto capi di imputazione giudiziaria), di un altro senatore dell'UDEUR, di altri due senatori Liberal Democratici e di Domenico Fisichella (ex monarchico, AN, Margherita, ma non entrato nel PD), Franco Turigliatto (ex Democrazia Proletaria, PRC) e Sergio De Gregorio (ex PSI, FI, Italia dei Valori), tutti eletti in liste della coalizione. Deflagrò così per la seconda volta il progetto ulivista di Romano Prodi. Il clima politico alla vigilia delle elezioni dell'Aprile 2008 vide pertanto la contrapposizione tra il fronte compatto del Centro Destra [ d'ora in avanti CDX ] e quello frantumato del CSX, in cui il PD di Walter Veltroni avrebbe dovuto tentare, da solo, di ricostruire una credibilità politica fortemente compromessa dalla instabilità interna alla variegata area di riferimento della sua ex maggioranza.

#### Obiettivo della ricerca.

Il quadro politico sopra illustrato suggerì l'idea di un'indagine volta a testare gli umori dell'elettorato alla vigilia del voto di primavera, al fine di prevedere quali forze attive nell'area bipolare avrebbero presumibilmente guadagnato o perso consensi.

#### Ambito, unità d'analisi e criteri di campionamento.

La coincidenza del periodo di somministrazione del questionario con il periodo pasquale (la festività cadde il 23 Marzo) suggerirono di utilizzare i giorni di sospensione della didattica come occasione per interviste territorialmente distribuite nelle zone di residenza degli studenti partecipanti al Laboratorio. I

numerosi “fuori sede” assicurarono così una variegata stratificazione degli intervistati, calmierando parzialmente lo squilibrio dovuto all’inevitabile concentrazione di bolognesi.

La selezione dei potenziali soggetti intervistabili (**unità d’analisi**) avvenne assicurando preliminarmente il possesso dei seguenti requisiti:

- aver partecipato alle precedenti elezioni della Camera nel 2006;
- la specifica di avere votato per un partito dell’area di CSX o CDX o di non aver votato (astendosi, votando scheda bianca o annullando il voto).

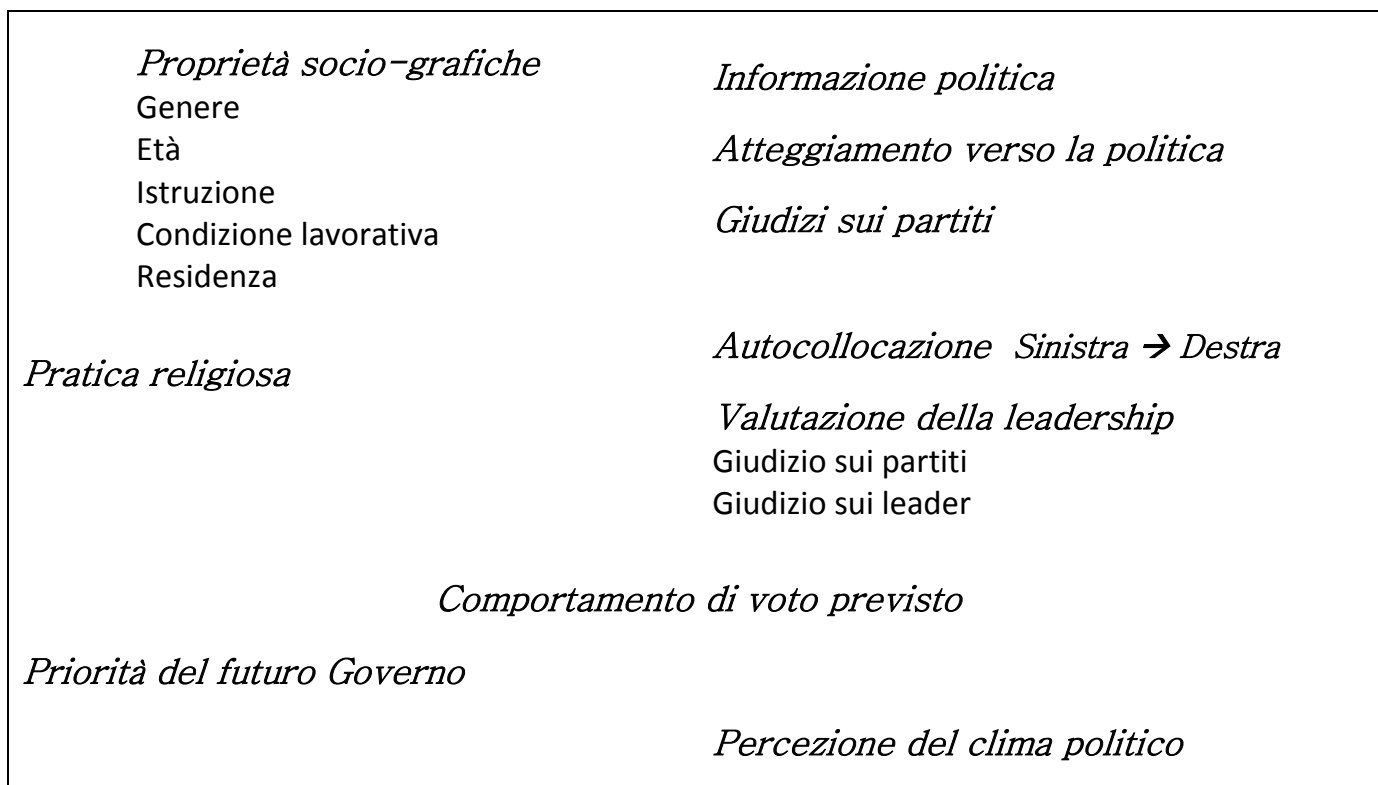
La scelta dei soggetti da intervistare (**casi della ricerca**) fu infine condizionata dall’obbiettivo di ottenere la seguente ripartizione: 40% di votanti partiti dell’area di CSX ; 40% di votanti partiti dell’area di CDX; un restante 20% di non votanti.

Il rispetto di tali criteri di selezione permise di preconstituire due serbatoi numericamente equivalenti di elettori del CSX e del CDX, e un serbatoio più piccolo riguardante l’astensionismo. In tal modo, al di là della rappresentatività statistica riferita alle reali quote di consensi registrate alla Camera nelle elezioni 2006 , vennero assicurate le condizioni per una comparazione degli orientamenti politici delle due aree di elettorato afferenti a CSX e CDX, unitamente alla possibilità di controllare gli umori di chi nel 2006 decise di astenersi dal voto.

*Da ora in avanti, per semplicità espositiva, verrà utilizzato il tempo presente.*

### **B.3. Mappa dei concetti.**

Lo studio dell’umore dell’elettorato e della previsione del suo comportamento di voto suggerisce la rilevazione di un insieme di proprietà, definibile come una sorta di Mappa dei concetti ritenuti congruenti con gli obbiettivi della ricerca. La Mappa è sintetizzata nel riquadro sottostante.



Essa si sostanzia di proprietà semplici (traducibili direttamente in una domanda a risposte chiuse o aperte mediante una definizione operativa univoca) o complesse (dove la proprietà necessita di due o

più indicatori, a loro volta aventi la capacità di essere singolarmente tradotti in una o più specifiche domande). Con la successiva definizione operativa, ogni proprietà verrà pertanto trasformata in uno strumento in grado di recepire, mediante la risposta data, lo stato specifico di ciascun intervistato sulla proprietà a cui la domanda fa riferimento. Pensando alla matrice dati “casi per variabili”, su ogni riga si troveranno tutti gli stati del soggetto registrati sulle specifiche proprietà, mentre i singoli stati osservati nei vettori colonna costituiranno le variabili.

#### B.4. Dalle proprietà alle variabili: le domande da inserire nel questionario.

Di seguito vengono illustrati gli esiti delle definizioni operative scelte relativamente alle proprietà incluse nella Mappa .

##### *Proprietà socio-grafiche*

Genere → *Genere* [ 1 ] donna [ 2 ] uomo

Età → *Anno di nascita:* 19\_\_\_\_/\_\_\_\_/

Istruzione → *Generico titolo conseguito:*

- 1  scuola dell'obbligo
- 2  diploma
- 3  laurea triennale
- 4  laurea specialistica (o vecchia quadriennale, o ingegneria / medicina)
- 5  master

Condizione lavorativa → *Posizione rispetto al lavoro:*

- 1  sono uno studente universitario
- 2  studio e lavoro
- 3  sono disoccupato
- 4  sono in cerca di prima occupazione
- 5  lavoro

Residenza → *Provincia di residenza (sigla automobilistica) : /\_\_\_\_\_/\_\_\_\_\_/*

##### *Pratica religiosa*

La proprietà è definita operativamente attraverso un unico indicatore: la *frequenza alla messa*.

*Escluse le cerimonie come matrimoni, funerali e battesimi, con quale frequenza va alla messa ?*

- |                            |                             |                            |                        |                            |                         |
|----------------------------|-----------------------------|----------------------------|------------------------|----------------------------|-------------------------|
| 6 <input type="checkbox"/> | Anche più volte a settimana | 5 <input type="checkbox"/> | Tutte le domeniche     | 4 <input type="checkbox"/> | Due - tre volte al mese |
| 3 <input type="checkbox"/> | Una volta al mese           | 2 <input type="checkbox"/> | Due – tre volte l’anno | 1 <input type="checkbox"/> | Praticamente mai        |

## Informazione politica

Questa proprietà complessa è definita utilizzando nove differenti indicatori: due, relativi alla frequenza di lettura della carta stampata; sette, riguardanti la visione di telegiornali.

La tecnica di *scaling* proposta su tutti gli indicatori è quella della scala auto – ancorante, in cui vengono esplicitati solamente gli estremi della dimensione su cui il soggetto deve collocarsi .

Dai media proposti sono esclusi *i social network* e *i blog*, all'epoca meno diffusi e in prevalenza utilizzati solo da una parte della popolazione giovanile e da una quota molto inferiore di adulti.

*Rispetto all'informazione politica, quanto frequentemente (da 1=mai o quasi mai a 5=molto spesso )*

Legge quotidiani (anche su internet)	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Legge riviste di attualità e politica	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Vede il TG 1	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Vede il TG 2	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Vede il TG 3	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Vede il TG 4	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Vede il TG 5	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Vede Studio Aperto	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Vede il TG di La 7	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>

## Atteggiamento verso la politica

La materializzazione di questo aspetto avviene mediante una tipologia compenetrante tre differenti criteri: interesse, partecipazione, competenza. Le modalità di risposta possibili si sostanziano pertanto del contributo di uno o più criteri. La domanda è la seguente:

*Quale di queste frasi esprime meglio il suo atteggiamento nei confronti della politica*

- 1  lo mi considero politicamente impegnato
- 2  lo mi tengo al corrente della politica ma senza parteciparvi personalmente
- 3  lo penso che bisogna lasciare la politica a persone che hanno più competenze di me
- 4  La politica mi disgusta

## Giudizi sui partiti

Vengono poste tre affermazioni, sulle quali si chiede di esprimere il grado di accordo mediante le canoniche modalità di risposta proprie della scala Likert:

1  per nulla      2  poco      3  abbastanza      4  molto

Le affermazioni sono:

- a. *I partiti, anche se commettono errori, sono strumenti indispensabili di partecipazione politica ed elementi fondamentali della vita democratica;*
- b. *I partiti svolgono un'azione politica sostanzialmente negativa. Hanno finito per anteporre i propri interessi particolari a quelli della collettività;*
- c. *Oggi in Italia non c'è più un partito che voglia realmente un cambiamento della società.*

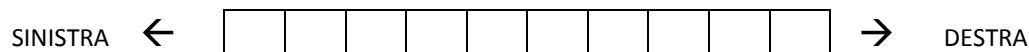
## Autocollocazione Sinistra → Destra

Questa dimensione è operativizzata proponendo il *continuum Sinistra → Destra*, nel formato della scala auto – ancorante con dieci caselle, in cui sono stati specificati solamente gli estremi.

L'assenza di una casella equidistante dai due estremi impedisce al rispondente di collocarsi "in mezzo", costringendolo pertanto ad una forzosa scelta di campo. Il criterio di somministrazione prevede in questo caso l'utilizzo diretto della domanda da parte dell'intervistato. Lui stesso apporrà la crocetta nella casella scelta come indicatore della propria auto collocazione.

*Personalmente, come si collocherebbe sul sottostante continuum politico sinistra - destra ?*

PORGERE IL QUESTIONARIO E CHIEDERE DI METTERE UNA X IN UNA CASELLA



NB Nello schema grafico di questa domanda mancano deliberatamente i riferimenti numerici corrispondenti al valore di ciascuna casella. Le frecce indicano infatti all'intervistato che i due estremi del *continuum* godono della medesima plausibilità. La presenza di una scala numerica orientata con valori crescenti da sinistra a destra (da 1 nella prima casella di sinistra fino a 10 nell'ultima casella di destra) potrebbe invece indurlo nel sospetto che il ricercatore tradisca l'imparzialità, associando ai valori più alti della scala la preferenza della dimensione di destra.

## Valutazione della leadership

Il concetto di *leadership* è immaginato come confronto tra il giudizio dato al partito e quello dato al suo *leader*. Per poter compiere tale confronto occorre pertanto rilevare il giudizio sui singoli partiti e sui *leader*. Si chiede all'intervistato di attribuire il giudizio utilizzando lo strumento del voto scolastico. Alla fine, la valutazione di ogni partito e di ogni leader produce  $13 \times 2 = 26$  variabili nella matrice dei dati.

*Ora leggerò i nomi di tredici partiti in lizza e quello del loro leader . In base al suo giudizio*

*Le chiederò di attribuire loro due voti scolastici, utilizzando la gamma dei punteggi da zero a dieci.*

partito	voto		leader	voto
POPOLO DELLA LIBERTA'		↔	BERLUSCONI	
LEGA NORD		↔	BOSSI	
MPA MOVIMENTO PER L'AUTONOMIA		↔	LOMBARDO	
LA DESTRA FIAMMA TRICOLORE		↔	D. SANTANCHE'	
LISTA ABORTO ? NO GRAZIE !		↔	FERRARA	
UDC		↔	CASINI	
LA ROSA BIANCA		↔	PEZZOTTA	
PARTITO DEMOCRATICO (con i radicali)		↔	WELTRONI	
ITALIA DEI VALORI – LISTA Di PIETRO		↔	Di PIETRO	
PARTITO SOCIALISTA		↔	BOSELLI	
SINISTRA ARCOBALENO		↔	BERTINOTTI	
PARTITO COMUNISTA DEI LAVORATORI		↔	FERRANDO	
SINISTRA CRITICA		↔	TURIGLIATTO	

## Comportamento di voto previsto

La scelta operata prevede di evitare imbarazzo all'intervistato. Ad un mese dalla consultazione elettorale, chiedere "per chi voterai?" potrebbe infatti indurre risposte forzate o poco attendibili. L'attenzione viene quindi posta sulla consapevolezza di come gli intervistati si pongono di fronte al voto imminente, se cioè hanno le idee chiare o sono incerti; contando di ricostruire l'appartenenza politica in base ai giudizi espressi sui singoli partiti e sui loro leader, alle informazioni sul partito votato nel 2006 e sull'autocollocazione sul *continuum Sinistra → Destra*.

Come si comporterà alle prossime elezioni politiche del 13 Aprile [LEGGI LE RISPOSTE]

1 <input type="checkbox"/> so già quale partito voterò	2 <input type="checkbox"/> so già l'area politica ma sono indeciso sul partito	3 <input type="checkbox"/> sono indeciso su tutto
4 <input type="checkbox"/> voterò scheda bianca o nulla o non andrò a votare		

## Priorità del futuro Governo

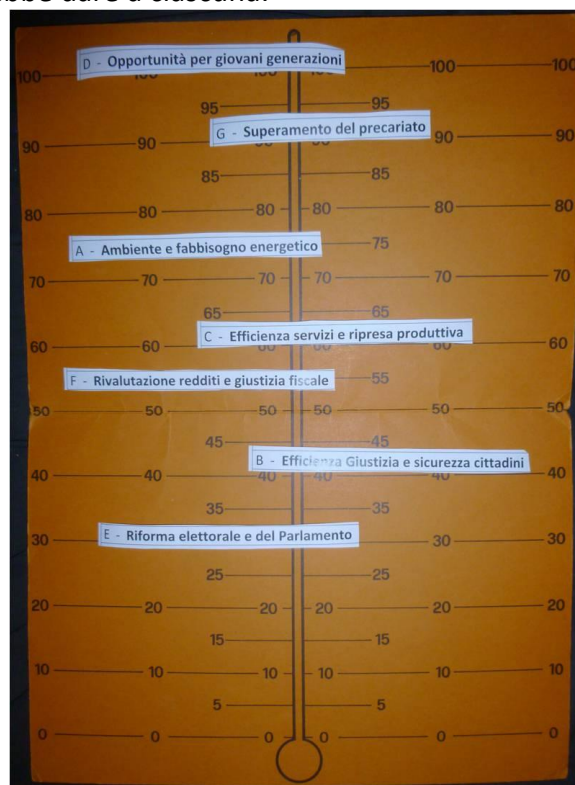
Sono sottoposte all'intervistato sette questioni di tipo economico, ambientale, istituzionale e sociale (vedi sotto la domanda). Utilizzando la tecnica di *scaling* del *termometro dei sentimenti*, qui sotto esemplificato, si chiede all'intervistato di collocare ogni questione su un grado del termometro rappresentante la *Priorità* che il futuro Governo dovrebbe dare a ciascuna.

Nell'esempio, l'ipotetico intervistato dispone liberamente i sette cartellini sul termometro, posizionandoli su differenti gradi, scelti fra i ventuno disponibili (il termometro segna infatti 5° per volta, da 0 a 100).

Da notare come questa tecnica permetta all'intervistato di esercitare il pieno controllo della situazione. Egli infatti:

- ragiona simultaneamente su tutte le questioni;
- le compara fra loro;
- le colloca sul termometro utilizzando un criterio di "taratura personale" che aumenta il livello di attendibilità delle sue risposte.

Operativamente, ultimata la disposizione dei cartellini, egli comunicherà all'intervistatore il valore del grado attribuito ad ogni questione. Nell'esempio fatto, la codifica è quella indicata in **corsivo** (75, 40, ..., 55, 90) nella colonna riepilogativa della domanda.



Vorrei ora sapere – secondo il suo parere – con quale grado di priorità il Governo formatosi dalle prossime elezioni dovrebbe affrontare le questioni - tutte importanti – di seguito elencate

[ PORGERE IL FOGLIO CON L'ELENCO, IL TERMOMETRO DI CARTONE CON I CARTELLINI E SPIEGARE IL MECCANISMO DI RISPOSTA ].

PRIORITA'	0 → 100
A - <b>Ambiente e fabbisogno energetico</b>	75
B - <b>Efficienza Giustizia e sicurezza cittadini</b>	40
C - <b>Efficienza servizi e ripresa produttiva</b>	60
D - <b>Opportunità per giovani generazioni</b>	100
E - <b>Riforma elettorale e del Parlamento</b>	30
F - <b>Rivalutazione redditi e giustizia fiscale</b>	55
G - <b>Superamento del precariato</b>	90

### *Percezione del clima politico*

Come ricordato nel paragrafo relativo alla situazione politica determinante le elezioni anticipate del 2008, l'elettorato dell'area di CSX era fortemente scosso dal secondo ribaltone ai danni di Romano Prodi e della sua prospettiva ulivista. Dall'altra sponda, Silvio Berlusconi preparava la riscossa, ignaro del fantastico incremento di consensi elettorali che avrebbe ottenuto, completato sul piano parlamentare dall'incetta di seggi prodotto dalla prima applicazione del *Porcellum* di Calderoli.

A chiusura dell'intervista viene pertanto chiesto all'intervistato di indicare il tipo di percezione del panorama politico.

Come ultima domanda, le chiedo di esprimere quale fra le **percezioni dell'odierno panorama politico** di seguito elencate è maggiormente vicino al suo sentire ...

<b>SVOLTA INNOVATIVA</b>	<input type="text" value="1"/>
<b>NORMALITA'</b>	<input type="text" value="2"/>
<b>VICOLO CIECO</b>	<input type="text" value="3"/>
<b>ULTIMA SPIAGGIA</b>	<input type="text" value="4"/>
<b>SCHIFO GENERALE</b>	<input type="text" value="5"/>

In conclusione, le domande commentate nel paragrafo diventano l'ossatura del questionario utilizzato nelle fasi di intervista e successiva codifica delle risposte nella matrice-dati.

Il file **Qst\_08.dpf** riporta fedelmente il testo del questionario.

La successione delle domande segue un criterio di coerenza tematica.

Ogni domanda è identificata dalla lettera D seguita dal numero d'ordine.

L'identificativo di ogni domanda ( D1, D2 ... D15) è stato scelto come nome delle corrispettive variabili del file archivio PSPP contenente la matrice-dati.



## SEZIONE C

Riassumendo quanto fatto sino ad ora:

- è stato installato il programma PSPP e ne sono state descritte le caratteristiche di funzionamento principali;
- sono stati scaricati, oltre al manuale *tutor*, la copia del questionario e il file PSPP contenente la matrice-dati;
- sono stati descritti gli obiettivi della ricerca e le caratteristiche delle variabili ricavabili dalle risposte alle domande del questionario.

D'ora in avanti, in occasione di ogni esercitazione, sarà utile tenere aperto anche il file riprodotto il questionario [ *Qst\_08.dpf* ], oppure una sua copia cartacea.

### C.1. Uno sguardo preliminare sul file *Mat\_pol2008.sav* .

Aprire il file, osservando i seguenti passi:

- cliccare sull'icona di collegamento →



- una volta apertosi il programma PSPP

- cliccare sulla voce *file* → *open* e

indicare il percorso dove è stato memorizzato *Mat\_pol2008.sav* ,

evidenziare e cliccare ancora su *open*

**Esito:** Il file si apre secondo la modalità di visualizzazioni dati.

Nella prima riga compaiono i nomi attribuiti a ciascuna variabile [ fa eccezione *Case* – in prima posizione -, automatismo del programma indicante il numero progressivo di ciascuna riga della matrice ]. La scelta è stata quella di mantenere il legame più stretto possibile con quanto indicato nel questionario.

Nel concreto, limitando l'attenzione ai primi quattro nomi e alle prime quattro domande, vediamo che:

D1 si riferisce alla **D1. DOMANDA FILTRO. Alle politiche del 2006, alla Camera lei ha votato per il CSX o per il CDX ?**

D2 si riferisce alla **D2. Per quale partito ha votato alle passate elezioni politiche?**

D3 si riferisce alla **D3. Genere**

D4 si riferisce alla **D4. Anno di nascita .**

In virtù di questo collegamento fra nomi delle variabili e domande del questionario e avendo sotto controllo i codici di risposta delle domande, è possibile interpretare qualche dato contenuto in matrice.

Ad es., il primo e il ventunesimo caso/riga della matrice indicheranno: due elettori maschi ( D3 = 2); il primo, elettore di area CSX ( D1 = 1 ), che ha votato PRC ( D2 = 7 ) ed è nato nel 1968 ( D4 ); il ventunesimo, elettore di area CDX ( D1 = 2 ), che ha votato AN ( D2 = 11 ) ed è nato nel 1945 ( D4 ).

La matrice-dati sarà tanto più grande in larghezza quanto più numerose saranno le variabili e tanto più lunga quanto più casi/righe conterrà. Mediante le freccette direzionali poste nell'angolo in basso a destra della finestra, è possibile spostarsi in lungo ( ↓ ) e in largo ( → ) sull'intera matrice.

Mat\_pol2008.sav [DataSet1] – PSPPIRE Data Editor

File Edit View Data Transform Analyze Utilities Windows Help

Case	D1	D2	D3	D4	D5	D6	D7	D8	D9a	D9b	D9c	D10	D11	quotid	riviste	TG1	TG2	TG3	TG4	TG5	Staperto	Tgla7	D13	PDI	Berlusconi	LEGA	Bossi	MPA
1	1	7	2	1968	1	2	5	2	3	4	4	3	5	5	3	2	1	5	1	3	2	3	3	0	0	0	0	0
2	2	10	2	1985	1	3	1	2	3	3	1	9	1	3	3	2	2	2	2	3	4	1	1	10	10	7	7	99
3	3	98	2	1965	1	2	5	2	2	4	3	2	1	3	2	2	2	2	1	1	1	2	5	3	3	2	4	3
4	2	10	1	1969	1	1	5	4	3	4	4	6	1	3	1	2	1	1	1	4	4	1	5	6	6	5	4	99
5	3	98	2	1976	2	2	1	4	3	4	3	4	5	4	2	1	1	1	1	1	1	4	5	5	0	0	2	99
6	2	10	2	1950	2	3	5	2	3	4	3	6	2	3	1	4	4	2	3	4	1	2	4	7	7	6	5	99
7	2	10	1	1961	2	2	5	3	2	3	2	7	4	3	1	3	3	1	3	1	2	1	1	8	8	6	5	6
8	1	1	2	1953	2	4	5	1	3	2	3	3	1	4	3	3	2	3	1	2	1	2	1	3	2	1	2	4
9	1	7	1	1960	2	4	5	2	4	3	3	2	1	4	2	3	4	2	1	2	1	1	2	4	5	3	3	3
10	1	1	1	1984	2	3	1	2	3	4	3	3	1	4	2	2	2	1	1	2	1	1	2	0	0	0	0	1
11	2	11	2	1980	2	4	5	2	2	4	1	10	1	2	1	3	2	2	2	5	2	1	5	6	6	7	7	99
12	2	10	2	1943	2	1	6	3	1	2	3	10	1	1	1	3	4	1	5	5	3	3	1	10	10	10	10	0
13	2	16	2	1965	2	2	5	3	3	4	2	10	2	5	2	2	1	4	2	5	4	3	1	5	5	6	6	99
14	1	1	1	1971	2	2	5	3	2	4	4	7	2	4	3	5	5	2	2	5	1	1	2	7	7	10	9	99
15	2	11	1	1973	45	5	5	2	2	3	3	6	2	5	2	3	3	2	1	4	2	1	2	7	6	5	5	6
16	3	98	2	1983	3	3	1	2	3	3	2	4	2	3	2	4	4	3	1	3	3	4	2	6	5	5	4	4
17	1	1	2	1945	3	2	6	2	3	3	3	3	4	2	2	5	1	2	1	2	1	2	1	2	2	0	0	0
18	1	1	1	1953	3	2	3	2	4	4	2	3	3	3	1	5	2	2	1	3	1	4	1	4	4	2	2	0
19	1	1	1	1981	3	2	5	3	4	3	4	4	1	3	1	2	3	1	1	3	4	1	3	3	3	2	2	99
20	1	2	2	1982	3	2	2	4	3	4	4	5	2	5	2	2	2	4	1	5	3	1	3	6	7	2	2	99
21	2	11	2	1945	3	2	6	3	3	3	2	9	5	2	4	5	5	2	4	5	2	3	1	8	8	5	5	5
22	2	10	1	1965	3	4	5	3	3	4	4	8	4	1	1	4	1	1	1	4	1	1	1	8	10	5	5	99

Volendo poi controllare le caratteristiche delle variabili occorre, come detto nelle sezioni precedenti, passare alla modalità di **visualizzazione variabili**. Cliccare pertanto in basso a sinistra su **Variable View**.

Mat\_pol2008.sav [DataSet1] – PSPPIRE Data Editor

File Edit View Data Transform Analyze Utilities Windows Help

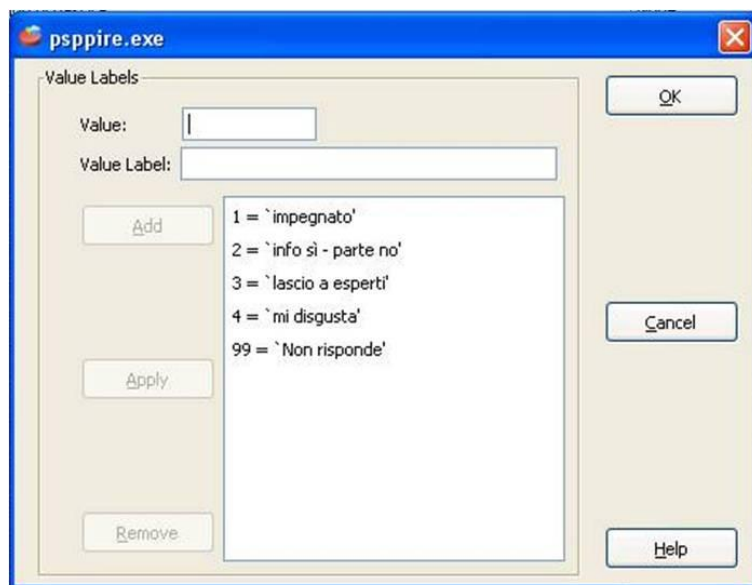
Variable	Name	Type	Width	Decimals	Label	Value Labels	Missing Values
1	D1	Numeric ...	1	0	FILTRO: alle politiche del 2006, alla Camera lei ha	{1, CSX}...	...
2	D2	Numeric ...	2	0	Partito votato nel 2006	{1, D5}...	... 98
3	D3	Numeric ...	1	0	Genere	{1, donna}...	... None
4	D4	Numeric ...	4	0	Anno di nascita	None	... None
5	D5	Numeric ...	2	0	Provincia di residenza	{1, AN}...	... None
6	D6	Numeric ...	1	0	Titolo di studio conseguito	{1, obbligo}...	... None
7	D7	Numeric ...	1	0	Posizione rispetto al lavoro	{1, stud univers}...	... None
8	D8	Numeric ...	2	0	Atteggiamento verso la politica	{1, impegnato}...	... 99
9	D9a	Numeric ...	2	0	Partiti indispensabili	{1, per nulla}...	... 99
10	D9b	Numeric ...	2	0	Partiti antepongono loro interessi	{1, per nulla}...	... 99
11	D9c	Numeric ...	2	0	Partiti non vogliono cambiamento	{1, per nulla}...	... 99
12	D10	Numeric ...	3	0	Continuum SX DX	{1, SX}...	... 99
13	D11	Numeric ...	2	0	Frequenza alla messa	{1, mai}...	... 99
14	quotid	Numeric ...	2	0	Frequenza lettura quotidiani	{1, mai o quasi}...	... 99
15	riviste	Numeric ...	2	0	Frequenza letture riviste di carattere socio-politico	{1, mai o quasi}...	... 99
16	TG1	Numeric ...	2	0	TG1, Frequenza ascolto	{1, mai o quasi}...	... 99
17	TG2	Numeric ...	2	0	TG2, Frequenza ascolto	{1, mai o quasi}...	... 99

Il registro, come anticipato nella Sezione **A**, è lo strumento necessario al programma per rendere gestibile ogni variabile memorizzata nella matrice dati, definendo nel dettaglio i criteri di interpretazione di quanto in essa contenuto. Gestire una variabile significa, oltre alla possibilità di “sapere” dove individuarla fra le colonne della matrice-dati, avere la possibilità di produrre output statistici corredati della massima informazione possibile ( in particolare attraverso l’impiego delle **Label** ).

A mo’ di esempio, si consideri l’*Atteggiamento verso la politica*. Come argomentato nel paragrafo **B.4.**, la definizione operativa di questa proprietà si è tradotta in una domanda con quattro differenti modalità di risposta. Nel questionario, la domanda è stata inserita come ottava. La variabile corrispondente nella matrice-dati è stata così specificata nel registro di PSPP :

- **Name:** *D8* , riprendente la sigla identificativa sul questionario prima del testo della domanda.
- **Type:** *numerica*, poiché le quattro modalità di risposta alternative sono state registrate con codici numerici da 1 a 4 , mentre l’eventuale mancata risposta è stata codificata con 99 ;
- **Width:** 2 , poiché due è il numero di spazi necessario per inserire un qualsiasi codice da 1 a 99;
- **Decimals:** 0 , poiché i codici utilizzati sono numeri interi;
- **Label:** *Comportamento prossime elezioni del 13 Aprile 2008* ;
- **Value Labels:** le *etichette* riferibili a ciascuna *modalità* di risposta, visualizzabili cliccando sui tre puntini ... a destra del campo .

**Esito:**



- **Missing Values:** 99 . Inserito precauzionalmente. Tale codice individuerebbe gli eventuali intervistati che si fossero rifiutati di rispondere.

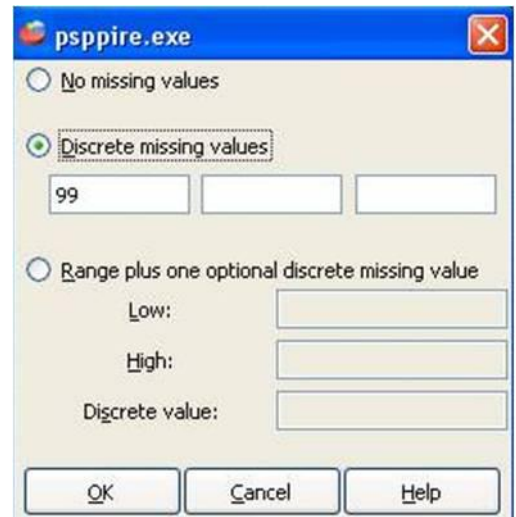
**NB.** PSPP consente di inserire sino a tre codici per classificare differenti modalità di non risposta. Ad es. , oltre al *Non risponde*, si potrebbero prevedere le modalità *Non so* e *Non applicabile*.

Desiderando attribuire i codici 98 al *Non so* e 97 al *Non applicabile*, si dovrebbe procedere come segue.

Cliccare sui tre puntini a destra ...

con l'Esito →

Si dovrebbero poi inserire i codici nelle tre location disponibili, quindi dare l' OK .



Riassumendo: PSPP, al fine di gestire le variabili, necessita di un *nome* univoco per ciascuna . Le *etichette* associabili alle variabili e alle loro modalità, così come i codici identificanti le *mancate risposte*, sono elementi opzionali. Tutto quanto è modificabile. In ogni caso va sottolineato che **le modifiche operate nel registro delle variabili o nella matrice dei dati diventano definitive esclusivamente se il file PSPP viene salvato. Il salvataggio può avvenire durante la sessione di lavoro o al termine della sessione stessa, contestualmente alla chiusura del programma.**

Come già visto, il salvataggio avviene cliccando sulla prima voce in alto a sinistra:

File, e selezionando la sotto-voce

Save oppure Save As ( se si vuole o meno mantenere lo stesso nome e lo stesso luogo di memorizzazione ).

**I preliminari sono terminati. Iniziano le esercitazioni guidate.**

Si da per scontato che sia attivo il file *Mat\_pol2008.sav* .

*Buon divertimento*

## C.2. ESERCIZIO n° 1 : analisi monovariata di alcune classificazioni.

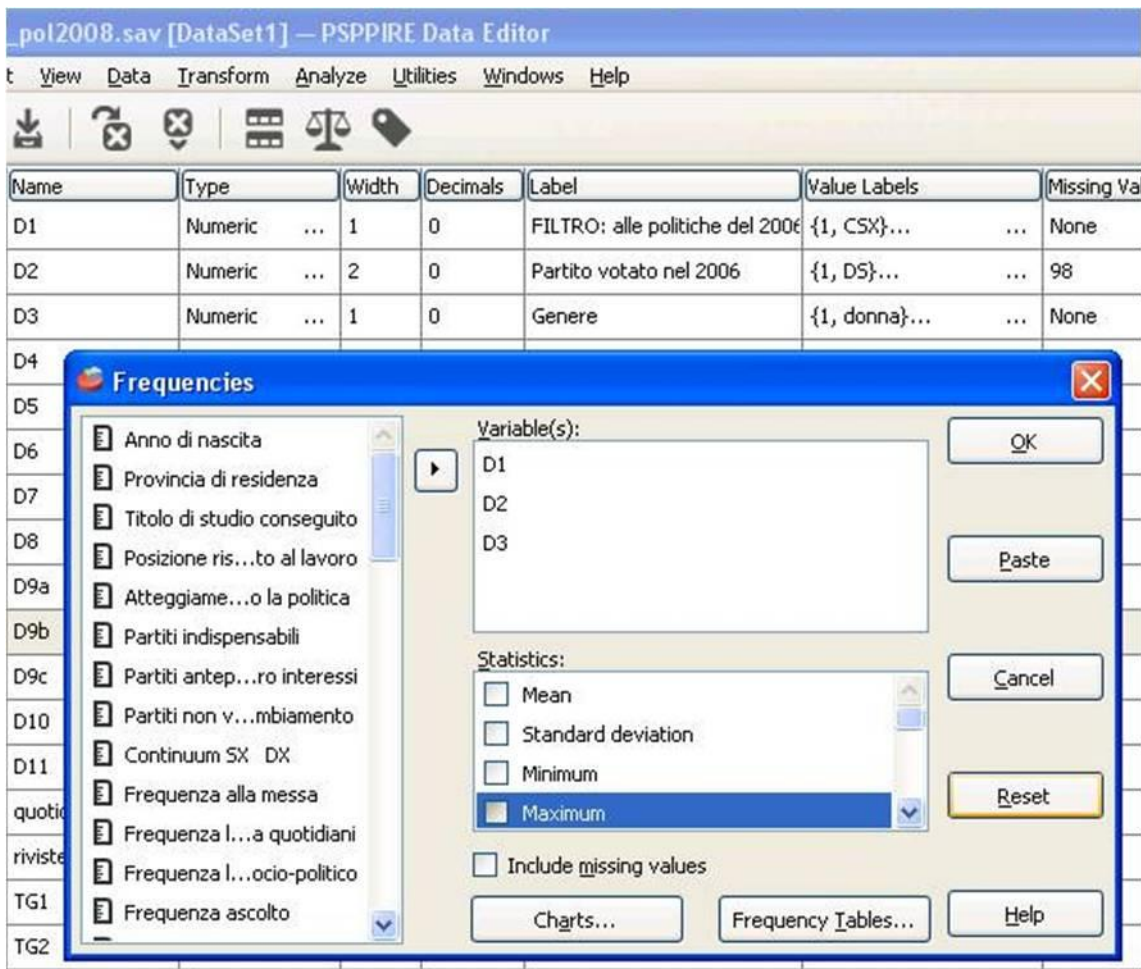
Produrre e commentare la distribuzione di frequenza delle variabili: D1 ( domanda filtro ), D2 ( partito votato alla Camera nel 2006 ), D3 ( sesso ).

*Esecuzione della procedura:*

- selezionare **Analyze** ( → ) **Descriptive Statistics** ( → ) **Frequencies**, quindi:
- evidenziare, nel campo a sinistra della sotto-finestra, le etichette delle prime tre variabili e dare un clic sulla freccetta ( → ) per spostarle nel campo a destra **Variable(s)**;
- deflaggare con un clic i **Check Mark** [ ✓ ] corrispondenti alle prime quattro **Statistics** proposte (\*).

(\* ) La decisione di escludere il calcolo delle quattro statistiche è dovuta al fatto che le variabili sono classificazioni in senso proprio, con modalità qualitative semanticamente autonome.

**Esito:**



La procedura è ora pronta per essere eseguita:

- cliccare sul bottone **OK**.

SPSS compie le operazioni richieste e trasferisce automaticamente il risultato nella finestra di output, come sotto illustrato.

Output — PSPPIRE Output Viewer

File Edit Windows Help

D1 + FREQUENCIES FREQUENCIES  
 D2 FREQUENCIES  
 D3 /VARIABLES= D1 D2 D3  
 D4 /FORMAT=AVALUE TABLE  
 D5 /STATISTICS=NONE.

FILTRO: alle politiche del 2006, alla Camera lei ha votato per il CSX o per il CDX ?

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
CSX	1	110	40,59	40,59	40,59
CDX	2	110	40,59	40,59	81,18
NO	3	51	18,82	18,82	100,00
<i>Total</i>		271	100,0	100,0	

Partito votato nel 2006

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
DS	1	60	22,14	27,27	27,27
MARGHERITA	2	14	5,17	6,36	33,64
ITA VALORI	4	4	1,48	1,82	35,45
SDI	5	1	,37	,45	35,91
PDCI	6	7	2,58	3,18	39,09
PRC	7	18	6,64	8,18	47,27
VERDI	8	3	1,11	1,36	48,64
RADICALI	9	3	1,11	1,36	50,00
FI	10	44	16,24	20,00	70,00
AN	11	41	15,13	18,64	88,64
LEGA	12	8	2,95	3,64	92,27
UDC	13	15	5,54	6,82	99,09
ALTRE DX	16	2	,74	,91	100,00
Non votato	98	51	18,82	Missing	
<i>Total</i>		271	100,0	100,0	

Genere

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
donna	1	113	41,70	41,70	41,70
uomo	2	158	58,30	58,30	100,00
<i>Total</i>		271	100,0	100,0	

TG1 Numeric ... | 2 | 0 | Frequenza ascolto | {1, ma o quasi}... ... | 99 | ... | 6

**Commento sulla variabile D1.** I conteggi dei casi per ogni categoria sono conformi agli obiettivi fissati in sede di campionamento [ B.2. ]: due serbatoi identici costituiti da 110 elettori ( nel 2006 ) dell'area di CSX e di CDX, di poco superiori al 40%, e un gruppo di controllo di 51 soggetti afferenti all'area del non voto, pari a circa il 19% .

**Commento sulla variabile D2.** La distribuzione di frequenza indica come la composizione interna alle due aree di CSX e CDX, viste nelle quote della colonna *Valid Percent* , sia leggermente squilibrata rispetto a quella uscita dalla consultazione elettorale. Impossibile, in sede di campionamento, riprodurre la marea di sigle scese in lizza.

In ogni caso, il numero di interviste raccolte permette di considerare separatamente le formazioni più grandi ( da una parte DS, MARGHERITA e PRC; dall'altra FI, AN e UDC ), suggerendo al contempo di raggruppare in un'unica categoria residuale le formazioni più piccole di ciascuna area. Questa considerazione troverà fra poco il suo risvolto operativo, introducendo la possibilità, con PSPP, di costruire la nuova variabile basata sull'aggregazione di alcune categorie della variabile D2 .

**Commento sulla variabile D3.** La quota di donne intervistate è di poco inferiore al 42% del totale e sottostima pertanto la reale distribuzione dell'elettorato per genere ( nel 2006 le donne aventi diritto al voto furono circa 22milioni contro i 20milioni di uomini ). In ogni caso, visto il carattere politologico della ricerca, tale squilibrio non dovrebbe inficiare più di tanto la verosimiglianza dei risultati, poiché è ormai noto come da molti anni donne e uomini tendano a differenziarsi solo su specifiche questioni di genere.

### C.2.1. Aggregazione delle categorie di una variabile: dalla D2 alla AggD2.

Come suggerito sopra, è utile includere Italia dei Valori, SDI, PDCI, VERDI e Radicali in una categoria residuale, che potrebbe essere etichettata come ALTRE SX . Dall'altra parte, causa la minore proliferazione di liste, si può pensare di associare alla LEGA ( otto casi ) i due intervistati che hanno votato ALTRE DX. L'operazione da compiere è quindi la creazione di una nuova variabile [ AggD2 ] ottenuta mediante l'aggregazione delle categorie di una variabile esistente [ D2 ]. Schematicamente:

PARTITI	Codice D2	Codice AggD2	Effetto nella AggD2
DS	1	1	invariato
MARGHERITA	2	2	invariato
ITA VALORI	4	4	invariato ma definente la nuova categoria ALTRO CSX
SDI	5	4	si aggrega in ALTRO CSX
PDCI	6	4	si aggrega in ALTRO CSX
PRC	7	3	Nuovo codice di PRC nella nuova variabile
VERDI	8	4	si aggrega in ALTRO CSX
RADICALI	9	4	si aggrega in ALTRO CSX
FI	10	10	Invariato
AN	11	11	invariato
LEGA	12	12	invariato
UDC	13	13	invariato
ALTRE DX	16	12	si aggrega in LEGA
Non votato	98 - missing	98 - missing	invariato

PSPP offre due modi di risolvere il problema formalizzato sopra: mediante il **Menu Transform** oppure utilizzando i comandi del linguaggio PSPP, scritti ed eseguiti all'interno di una finestra di Sintassi.

### C.2.2. Aggregazione delle categorie mediante *Recode into Different Variables* .

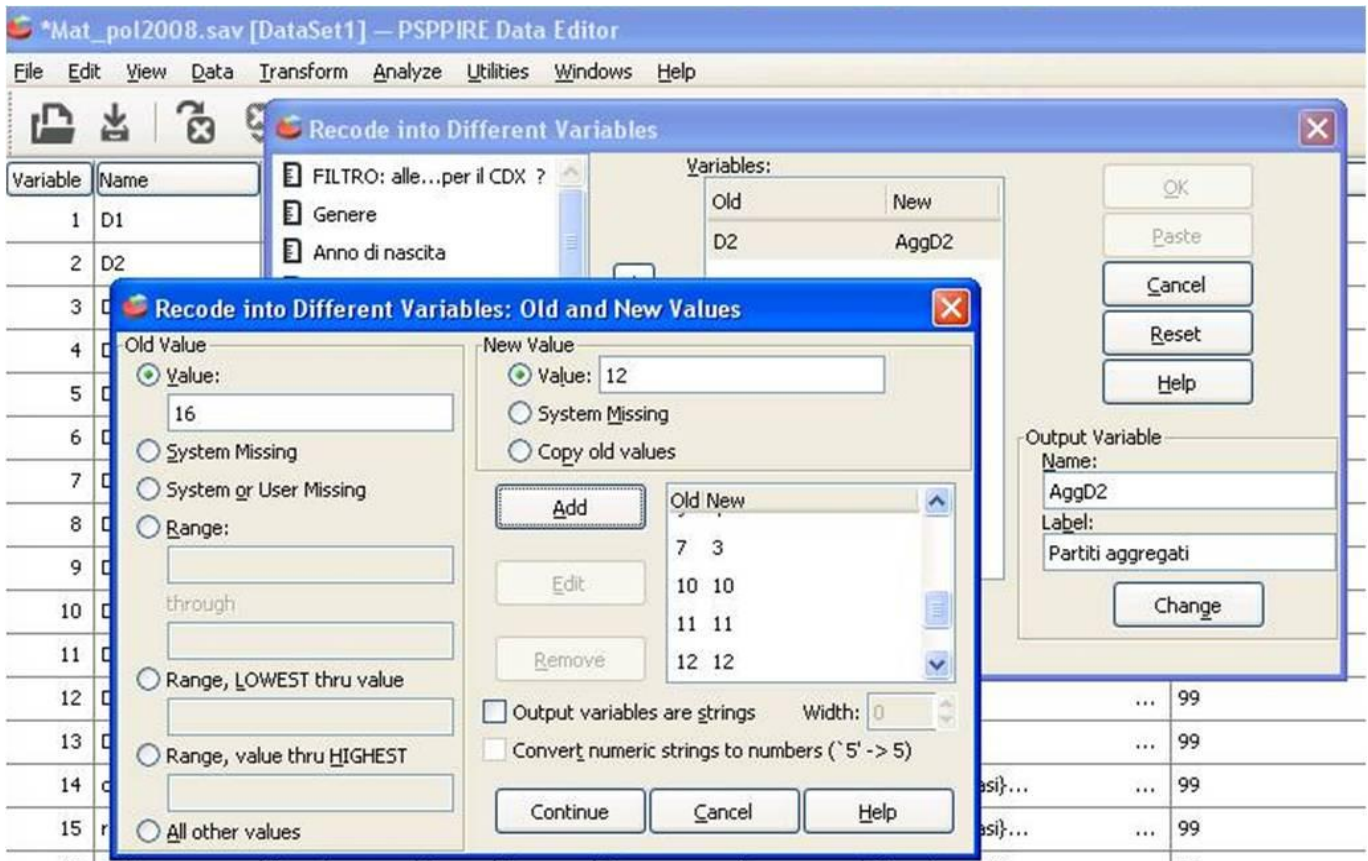
L'utilizzo del **Menu Transform** prevede i seguenti passaggi:

- selezionare **Transform ( → ) Recode into Different Variables, Esito:**



Occorre da qui procedere nel modo seguente [ **finestra in secondo piano** ]:

- dare un clic sull'etichetta *Partito votato nel 2006* e
  - trasferire la variabile D2 nella location **Variables:** [ comparirà al di sotto della voce Old ];
- dare un clic sul nome D2 → si rendono attive la voce *New* e le location *Output Variable*;
- ci si sposta nelle *location Output Variable* [ in basso a destra ] e si inseriscono il nome della nuova variabile [ AggD2 ] e la relativa etichetta [ *Partiti aggregati* ] ;
- cliccare sul bottone **Change** → si crea così la nuova variabile AggD2 su cui riversare le aggregazioni.



A questo punto si procede alla attribuzione e/o aggregazione delle categorie secondo lo schema logico illustrato nel paragrafo precedente. Questi i passi [ esiti: **finestra in primo piano** ]:

- cliccare sul bottone **Old and New Values ...** [ prima finestra a inizio paragrafo ];
- inserire nella *location Value* di *Old Value* il primo codice della D2 ( 1 , relativo ai DS ) e spostarsi in quella dei *New Value* per attribuire il nuovo codice nella variabile AggD2. La decisione è di mantenere lo stesso codice: quindi si inserisce 1 e si clicca sul bottone **Add** → il codice vecchio e quello nuovo [ 1 1 ] vengono posizionati nella finestra attiva *Old New* .
- si procede analogamente per ognuno dei codici dei partiti che rimangono inalterati, mentre per quelli da aggregare o modificare si cambia il codice nella *New Value* [ così il PRC da 7 diventa 3, e gli altri piccoli partiti di CSX prendono il codice 4; mentre il 16 di ALTRO DX diventa 12 , quello della LEGA ];
- terminate le ricodifiche si clicca sul bottone *Continue* e poi su **OK** → la variabile AggD2 è ultimata. PSPP aggiornerà il registro, collocando AggD2 in coda all'elenco delle variabili. Analogamente, il vettore dati sarà posto in ultima posizione nella matrice.

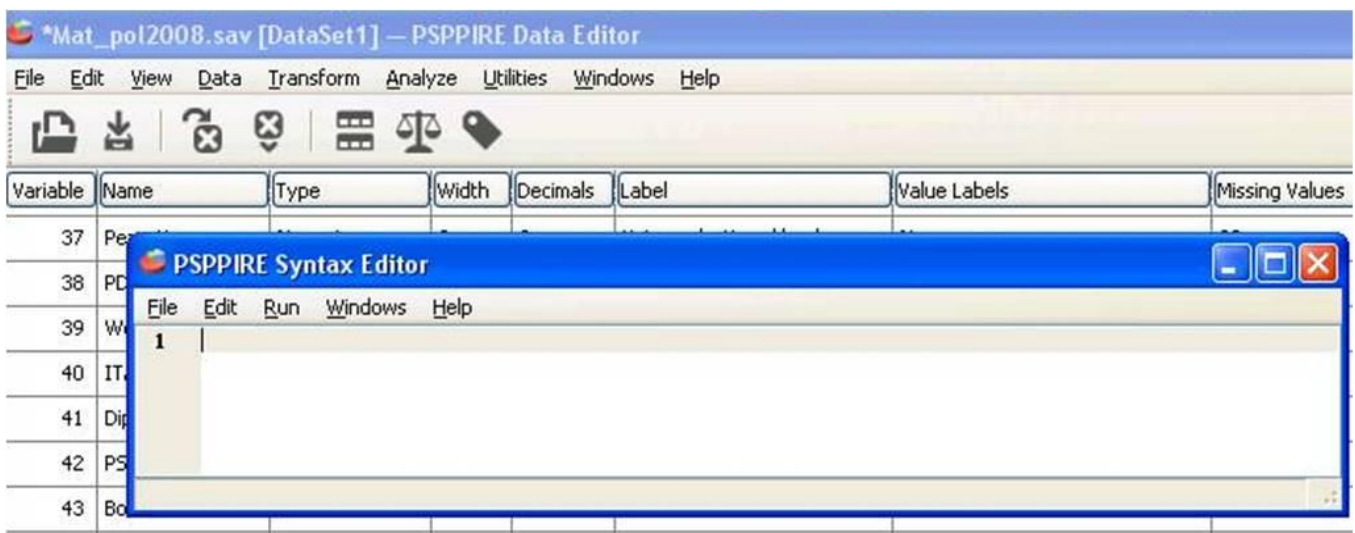


Va da sé che nella nuova variabile AggD2 manchino le etichette dei partiti. Sarà pertanto opportuno inserirle con il consueto meccanismo delle *Value Labels*. Anche il codice 98, ripetuto per coloro che si erano astenuti o avevano annullato il voto, andrà ridefinito come mancante mediante l'apposita funzione *Missing Values*. A modifiche ultimate, si provvederà a salvare permanentemente la nuova variabile cliccando su **File (→) Save**.

### C.2.3. Aggregazione delle categorie mediante la Sintassi di PSPP.

Come già detto, lo stesso processo di attribuzione e/o aggregazione delle categorie dalla variabile originaria D2 a quella nuova AggD2 può essere ordinato al programma mediante gli opportuni comandi, da scrivere e far eseguire all'interno di una finestra di Sintassi. Ciò vale non solo per questo tipo di intervento ma anche in tutti gli altri casi di *data management*, come ad esempio la duplicazione di variabili, la costruzione di indici additivi o di tipologie risultanti dalla combinazione di più modalità logicamente associate fra loro. Rimanendo tuttavia sull'esercizio in corso, si proceda nel seguente modo:

- aprire una finestra di Sintassi cliccando su **File (→) New (→) Syntax**. **Esito:**



**NB.** In presenza di un file di Sintassi pre-esistente, si sarebbe potuto aprirlo (evitando di crearne uno nuovo) semplicemente impostando **File (→) Open (→) Syntax**, indicando contestualmente il nome del file e il luogo in cui fosse stato memorizzato dall'utente.

A questo punto si può scrivere il comando, partendo da dove è posizionato il cursore.

```
RECODE D2 (98=98)(1=1)(2=2)(4=4)(5=4)(6=4)(8=4)(9=4)(7=3)(10 =10)(11=11)(12=12)(13=13)(16=12)
INTO AggD2 .
```

La logica della sintassi di PSPP è molto intuitiva.

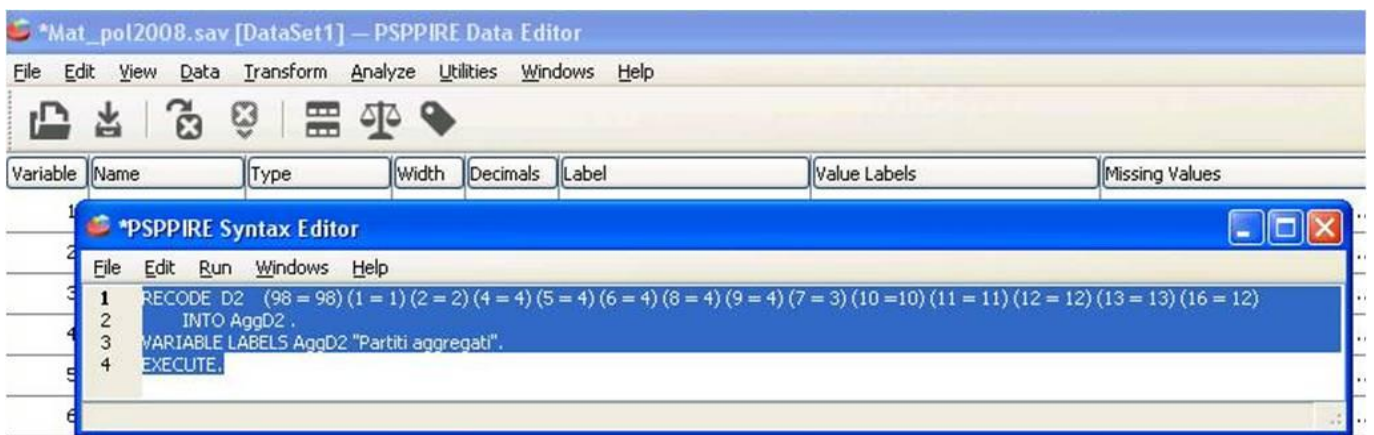
- Innanzitutto occorre indicare quale trasformazione si intende operare. Il primo elemento è pertanto il nome del comando: **RECODE** ;
- segue il nome della variabile sui cui si desidera aggregare alcune o tutte le categorie: **D2** ;
- fra campi di parentesi, si indicano poi come i singoli valori numerici della D2 vanno mantenuti o modificati nella nuova variabile aggregata ( $n_{vecchio} = n_{nuovo}$ );
- al termine delle ricodifiche si inserisce la parola chiave **INTO**, cosicché le modifiche apportate vengano applicate su una variabile differente da quella originale, di cui bisogna esplicitare il nome: **AggD2** .
- il comando termina con un punto .

Nel Menu grafico del **Recode into Different Variables** è stata anche associata al nome AggD2 la relativa etichetta. Con la sintassi, questa operazione si sostanzia nel comando:

VARIABLE LABELS AggD2 "Partiti aggregati".

Anche in questo caso la logica è la stessa: viene scritto per primo il termine identificante il comando ( VARIABLE LABELS ), quindi il nome della variabile su cui il comando deve agire ( AggD2 ) e infine, compresa tra doppie virgolette, l'etichetta desiderata ( Partiti aggregati ), ancora una volta seguita dal punto indicante la fine del comando.

L'esecuzione di entrambi i comandi ( RECODE e VARIABLE LABELS ) è possibile inserendo il comando operativo EXECUTE. In pratica, tutti i comandi, compreso EXECUTE, andranno evidenziati con il mouse e fatti eseguire cliccando sulle voci Run ( → ) Selection .



Una volta corredata la nuova AggD2 delle etichette riferite ai codici dei partiti, la distribuzione di frequenza [ Analyze ( → ) Descriptive Statistics ( → ) Frequencies ] darà questo risultato:

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
DS	1	60	22,14	27,27	27,27
MARGHERITA	2	14	5,17	6,36	33,64
PRC	3	18	6,64	8,18	41,82
ALTRO CSX	4	18	6,64	8,18	50,00
FI	10	44	16,24	20,00	70,00
AN	11	41	15,13	18,64	88,64
LEGA	12	10	3,69	4,55	93,18
UDEUR	13	15	5,54	6,82	100,00
Non voto	98	51	18,82	Missing	
<i>Total</i>		271	100,0	100,0	

L'effetto dell'aggregazione fa sì che la variabile sia utilizzabile nella sua interezza, prevedendo quattro categorie per ognuna delle due aree politiche.

### C.3. ESERCIZIO n° 2 : analisi monovariata di alcune scale.

Produrre i valori medi e di variabilità delle variabili: D4 ( anno di nascita ),TG1, TG2, TG3, TG4, TG5, Staperto, TGla7 ( frequenze di ascolto graduate secondo cinque stati di una scala auto - ancorante) .

Esecuzione della procedura:

- selezionare **Analyze** ( → ) **Descriptive Statistics** ( → ) **Descriptives**, quindi:

- evidenziare con il mouse le etichette delle variabili interessate e

- cliccare sulla freccetta ( → ) per spostarle nel campo a destra **Variable(s)**;

- lasciare i *Check Mark* [ ✓ ] previsti in automatico dalla procedura;

- cliccare sul bottone **OK** per avviare l'esecuzione del comando.

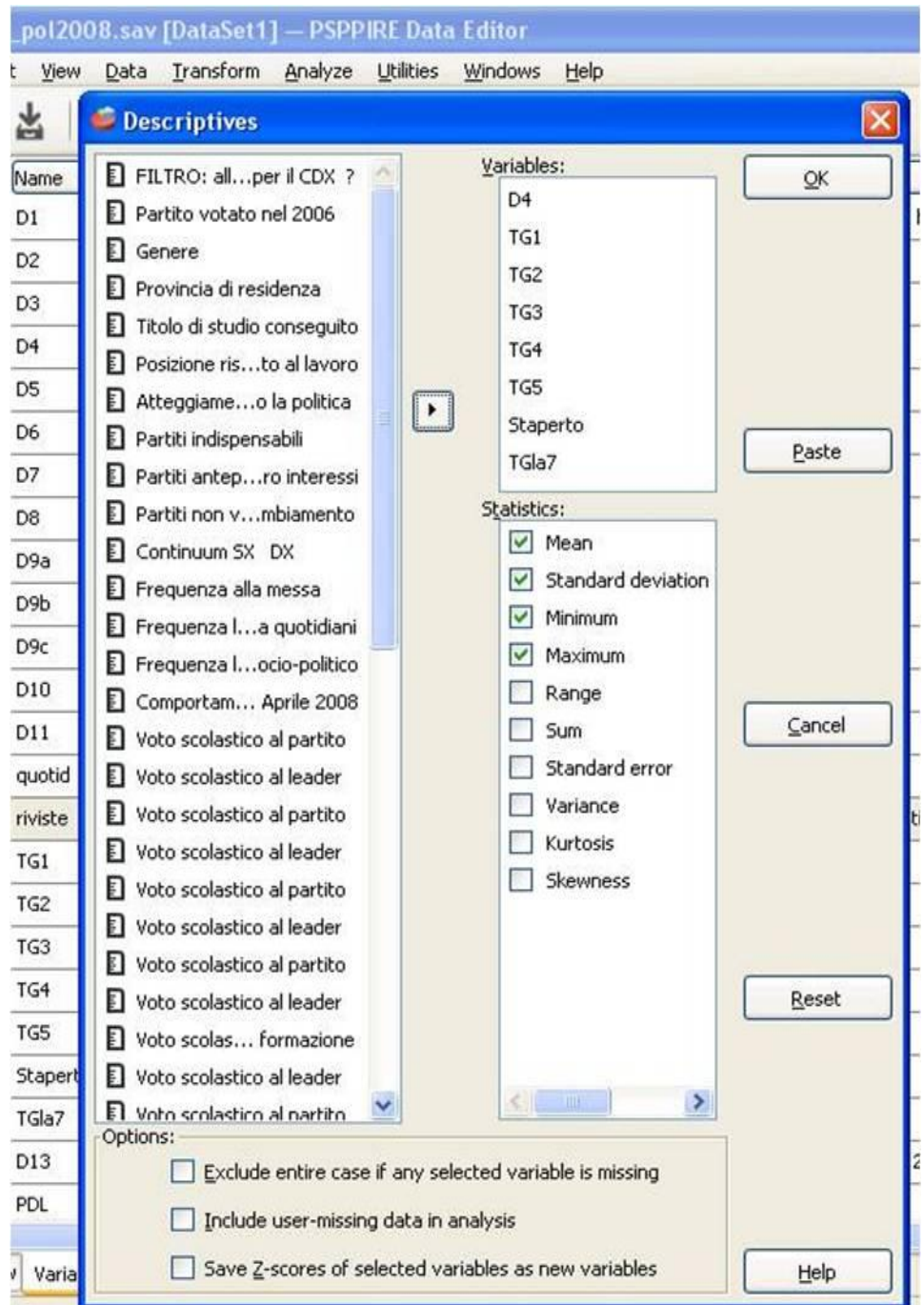
#### NB.

Il riquadro inferiore Options permette opzionalmente di:

escludere ogni caso abbia un valore mancante su una qualsiasi delle variabili;

includere come validi i codici definiti come mancanti;

salvare in nuove variabili i punteggi standardizzati delle variabili selezionate



La tabellina prodotta dal programma indica, a partire da sinistra:

- le etichette delle variabili, nell'ordine in cui sono state selezionate dall'utente;
- il numero totale di casi su cui sono stati effettuati i calcoli;

- Il valore della Media per ciascuna variabile;
- il valore dello Scarto tipo per ciascuna variabile;
- il valore Minimo e Massimo riscontrati almeno una volta su ciascuna variabile.

pol2008.sav [DataSet1] – PSPP Data Editor

File Edit Windows Help

DESCRIPTIVES

DESCRIPTIVES  
/VARIABLES= D4 TG1 TG2 TG3 TG4 TG5 Staperto TGla7.

Valid cases = 271; cases with missing value(s) = 0.

Variable	N	Mean	Std Dev	Minimum	Maximum
Anno di nascita	271	1970,30	14,31	1921,00	1987,00
TG1, Frequenza ascolto	271	3,12	1,31	1,00	5,00
TG2, Frequenza ascolto	271	2,55	1,25	1,00	5,00
TG3, Frequenza ascolto	271	2,40	1,37	1,00	5,00
TG4, Frequenza ascolto	271	1,71	1,04	1,00	5,00
TG5, Frequenza ascolto	271	2,92	1,44	1,00	5,00
Studio Aperto, Frequenza ascolto	271	2,20	1,30	1,00	5,00
TG la 7, Frequenza ascolto	271	1,60	1,02	1,00	5,00

**Commento sulle variabili relative all’ascolto dei TG.** Posto che sui sette telegiornali sono stati conteggiati soggetti lungo tutto la *range* della scala, si può notare come, all’epoca, gli ascolti maggiori si concentrassero su TG1, TG2 e TG5, con medie al di sopra di quella teorica (di 2,5 metà del *range* 1→5) mentre La 7 fosse ben lontana dagli *score* maturati negli anni a venire. Si nota anche un livello di ascolto medio abbastanza elevato del TG3, storico telegiornale di bandiera della sinistra, a indicare presumibilmente un certo “radicalismo militante” di una parte consistente degli intervistati. Ultima considerazione: i dati indicano i valori medi e di variabilità su ciascun TG ma nulla dicono sull’eventuale rapporto esistente negli ascolti congiunti di più emittenti. Questa curiosità potrà essere soddisfatta utilizzando le opzioni statistiche offerte dall’analisi bivariata [ come si vedrà nel capitolo C.5.2. ].

**Commento sulla variabile D4.** La media dell’anno di nascita risulta di poco superiore a 1970. Essendo stato effettuato il sondaggio nel 2008, ciò significa che l’età media degli intervistati era di poco inferiore ai 38 anni. Da questo ragionamento appare evidente come l’anno di nascita, rispetto agli anni, sia meno immediato da utilizzare nella descrizione dell’età della popolazione. L’anno di nascita, per quanto utile ai fini della registrazione dell’informazione in sede di intervista (specialmente in presenza di signore non più giovanissime), rimane quindi più difficoltoso da gestire in sede di analisi. Sarebbe allora utile **trasformare gli anni di nascita in anni di età**. La soluzione di questo problema sarà l’oggetto del prossimo paragrafo.

### C.3.1. Costruzione matematica di una nuova variabile: dalla D4 alla *Anni*.

Come suggerito sopra, l’obiettivo da conseguire è quello di trasformare l’anno di nascita dichiarato nel corrispondente numero di anni dell’intervistato. Si tratta pertanto di applicare a tutti i casi della ricerca (righe della matrice dati) un elementare algoritmo aritmetico che produca l’effetto desiderato. Tale effetto sarà registrato in una nuova variabile, così da mantenere inalterata la variabile originale.

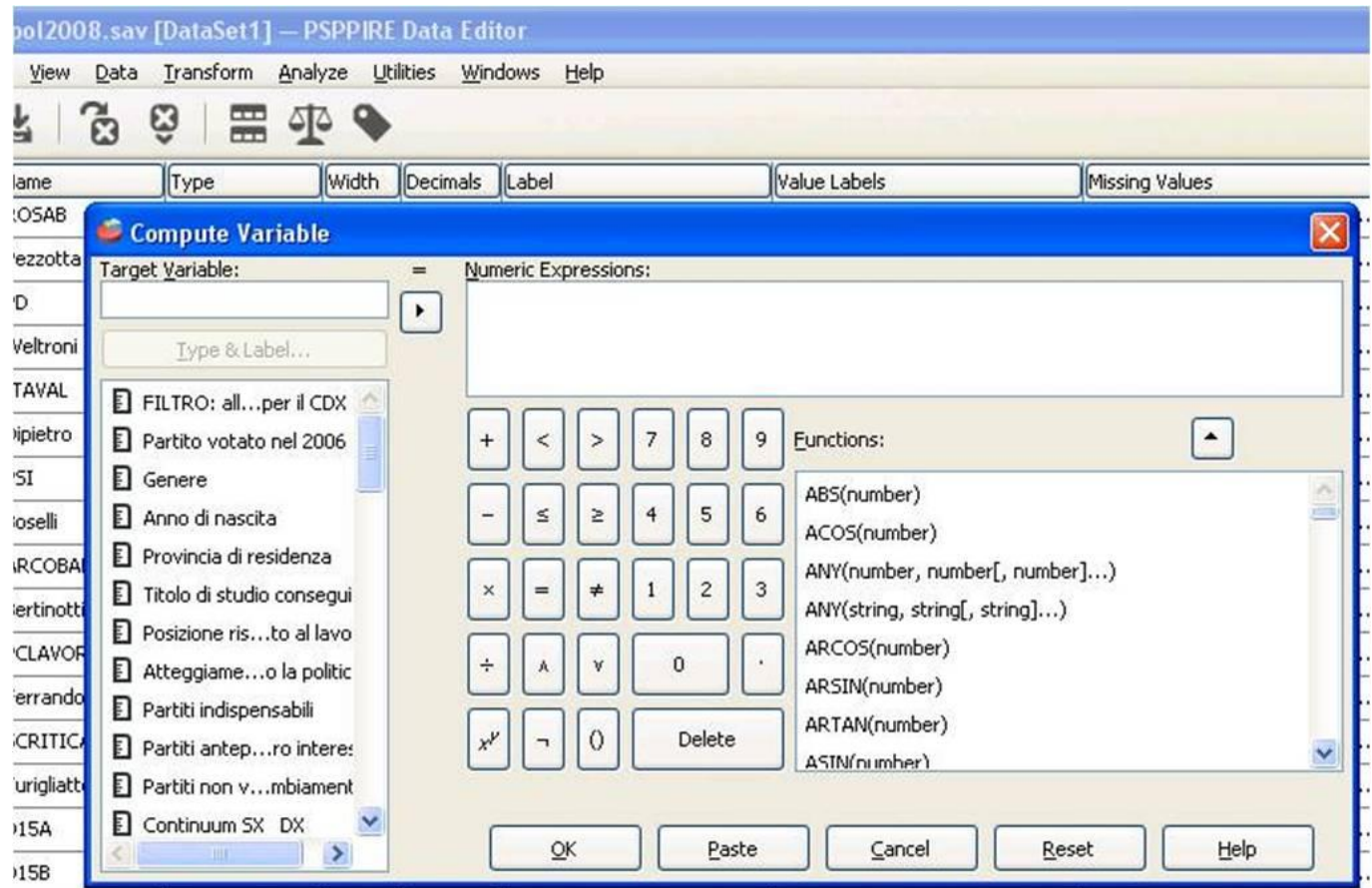
Posto che l'indagine campionaria è stata effettuata nel 2008, si prenderà questo anno come valore costante, sottraendo ad esso, caso per caso, l'anno di nascita dichiarato. L'algoritmo è pertanto:  $2008 - \text{anno di nascita}$ . Di seguito è illustrata la procedura da seguire mediante il **Menu Transform**.

### C.3.2. Costruzione matematica di una nuova variabile mediante **Compute**.

L'utilizzo del **Menu Transform** prevede i seguenti passaggi:

- selezionare **Transform ( → ) Compute**

**Esito:**



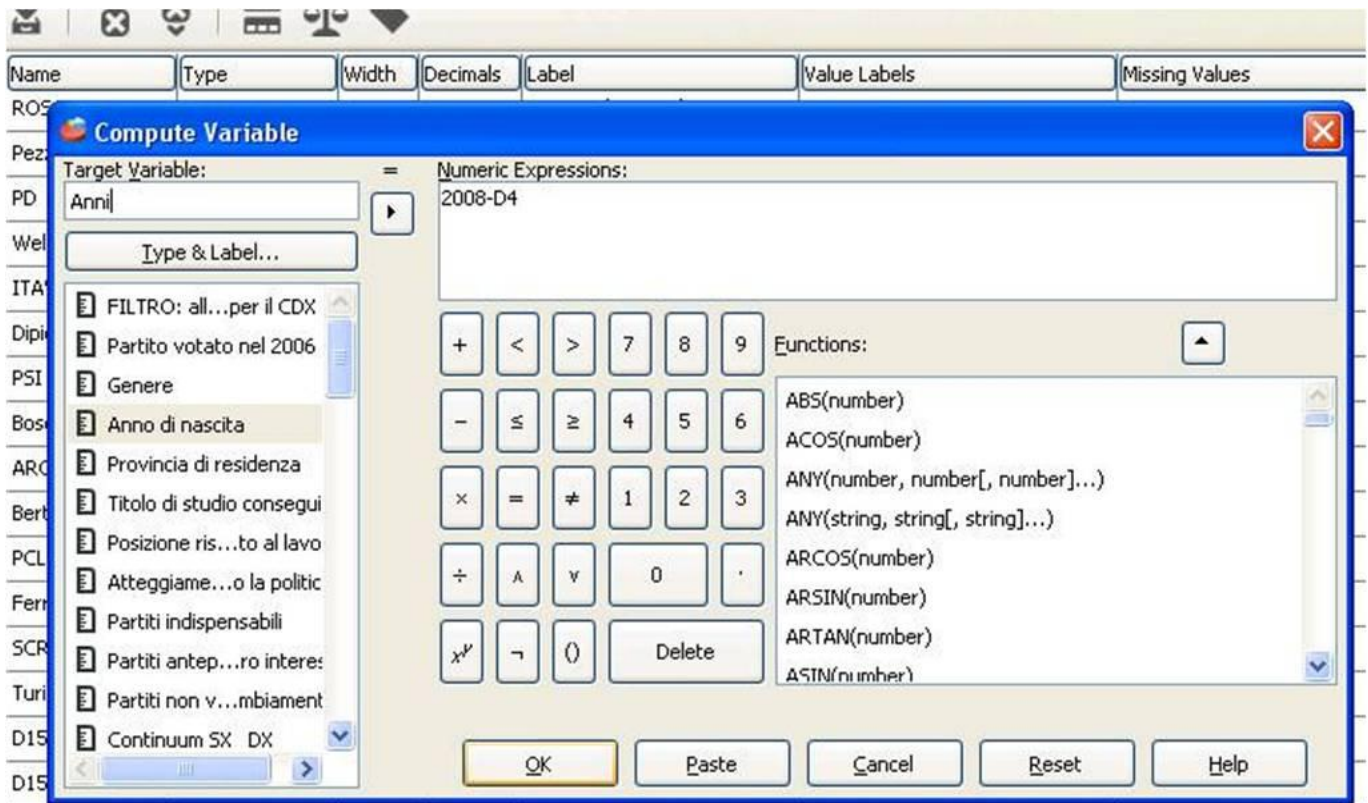
A questo punto occorre rispettare cronologicamente i seguenti passi [ vedi finestra successiva ]:

- puntare il mouse e dare un clic nell'angolo in alto a sinistra della **location Numeric Expressions**:

Quindi, utilizzando il tastierino multifunzione proposto al di sotto della **location**:

- inserire il numero **2008** ;
- cliccare sul tasto dell'operatore aritmetico **-** ;
- cliccare col mouse sull'etichetta **Anno di nascita** ( nel riquadro di sinistra );
- spostare la variabile **D4** nella **location Numeric Expressions**, dopo il segno del meno, cliccando sulla freccetta ( → );
- spostarsi con il mouse e dare un clic nell'angolo a sinistra della **location Target Variable**:
- scrivere il nome della nuova variabile [ **Anni** ] da creare in base all'algoritmo  $2008 - D4$  ;

- cliccare sul bottone **OK** per ottenere l'esecuzione del comando.



Viene così creata la nuova variabile *Anni*. Nel registro essa occuperà l'ultima riga. Sarà inutile mettere la **Label** e men che meno le **Value Label**, poiché il nome della variabile e la sua natura di scala enumerabile dice già quanto basta sapere. Ad essere perfezionisti, si potranno portare a 2 e 0 i valori di *Width* e *Decimals*. La variabile *Anni* occuperà l'ultima colonna a destra della matrice-dati, subito dopo la *AggD2* creata precedentemente.

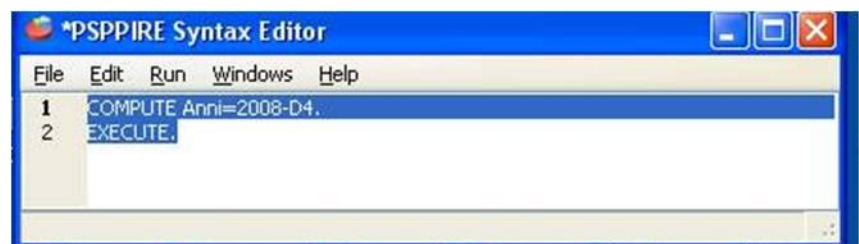
### C.3.3. Costruzione matematica di una nuova variabile mediante la Sintassi di PSPP.

Si osservi ora come il procedimento descritto nel paragrafo precedente possa essere semplicemente ottenuto attraverso l'uso dello specifico comando **COMPUTE** nella finestra di Sintassi. Sarà sufficiente:

- aprire una finestra di Sintassi cliccando su **File** (→) **New** (→) **Syntax** e al suo interno :
- scrivere il comando `COMPUTE Anni=2008-D4.` e, nella seconda riga, `EXECUTE.;`
- evidenziare le due righe ed eseguirle con **Run** (→) **Selection**

L'esito del comando sarà del tutto identico a quello visto prima.

Inutile sottolineare la praticità di tale soluzione.



Si ripeta ora la richiesta del calcolo dei valori medi e di variabilità sulla nuova variabile *Anni*. Dunque:

- **Analyze** ( → ) **Descriptive Statistics** ( → ) **Descriptives** ;

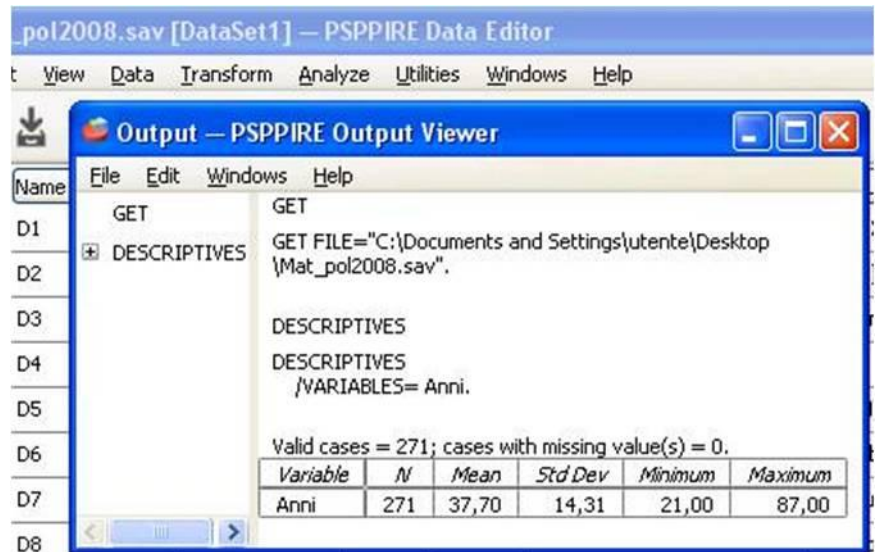
- selezionare *Anni* dalla *location Descriptives* e inserirla nella *Variables* mediante la freccetta ( → ) ;

- cliccare quindi sul bottone **OK** per ottenere il risultato voluto.

**Esito:**

L'output della variabile *Anni* appare di più facile comprensione anche se, nella sostanza, nulla è cambiato rispetto alla D4, ad eccezione della percezione diretta dovuta al valore della Media in *anni di età*.

Il *range* di 66 anni è egualmente ricavabile, facendo la differenza tra il valore massimo e quello minimo riscontrati. Infine, lo Scarto Tipo, ottenuto mediante una doppia normalizzazione, ha logicamente lo stesso valore di 14,31 .



The screenshot shows the PSPP Output Viewer window with the following content:

Variable	N	Mean	Std Dev	Minimum	Maximum
Anni	271	37,70	14,31	21,00	87,00

Additional text in the output window includes: "Valid cases = 271; cases with missing value(s) = 0.", "DESCRIPTIVES", and "/VARIABLES= Anni."

Anche in questo caso, per salvare permanentemente la nuova variabile, sarà sufficiente cliccare su

- **File** ( → ) **Save** .

#### C.4. ESERCIZIO n° 3 : analisi monovariata di una scala mediante *Frequencies* e grafica.

Nel paragrafo C.3. sono stati calcolati i valori medi e di variabilità di una scala enumerabile ( D4 poi trasformata in *Anni* ) e di sette scale auto ancoranti ( le frequenze di ascolto a telegiornali ). Gli stessi valori medi si sarebbero potuti ottenere mediante la procedura *Frequencies* , utilizzata nel suo automatismo prevedente i *Check Mark* [ ✓ ] attivi sul calcolo di Media, Scarto Tipo, Minimo e Massimo. Oltre a ciò, la stessa procedura *Frequencies* avrebbe consentito di ottenere alcuni tipi di grafici ( di scarsa qualità ma comunque esplicativi ).

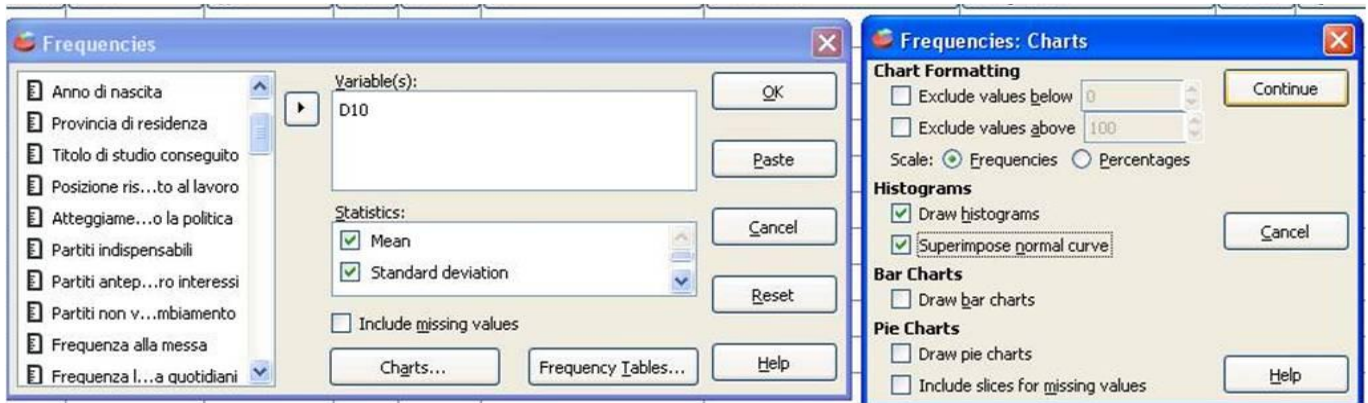
Per sperimentare queste possibilità offerte da PSPP, si analizzi la distribuzione di frequenza della variabile D10 ( *continuum* Sinistra → Destra ), compreso il calcolo dei valori medi e la produzione di un istogramma con approssimazione della curva normale.

Si proceda nel seguente modo:

- selezionare **Analyze** ( → ) **Descriptive Statistics** ( → ) **Frequencies** , quindi:

- evidenziare l'etichetta *Continuum SX DX* e inserire ( → ) D10 nella *location Variables* ;

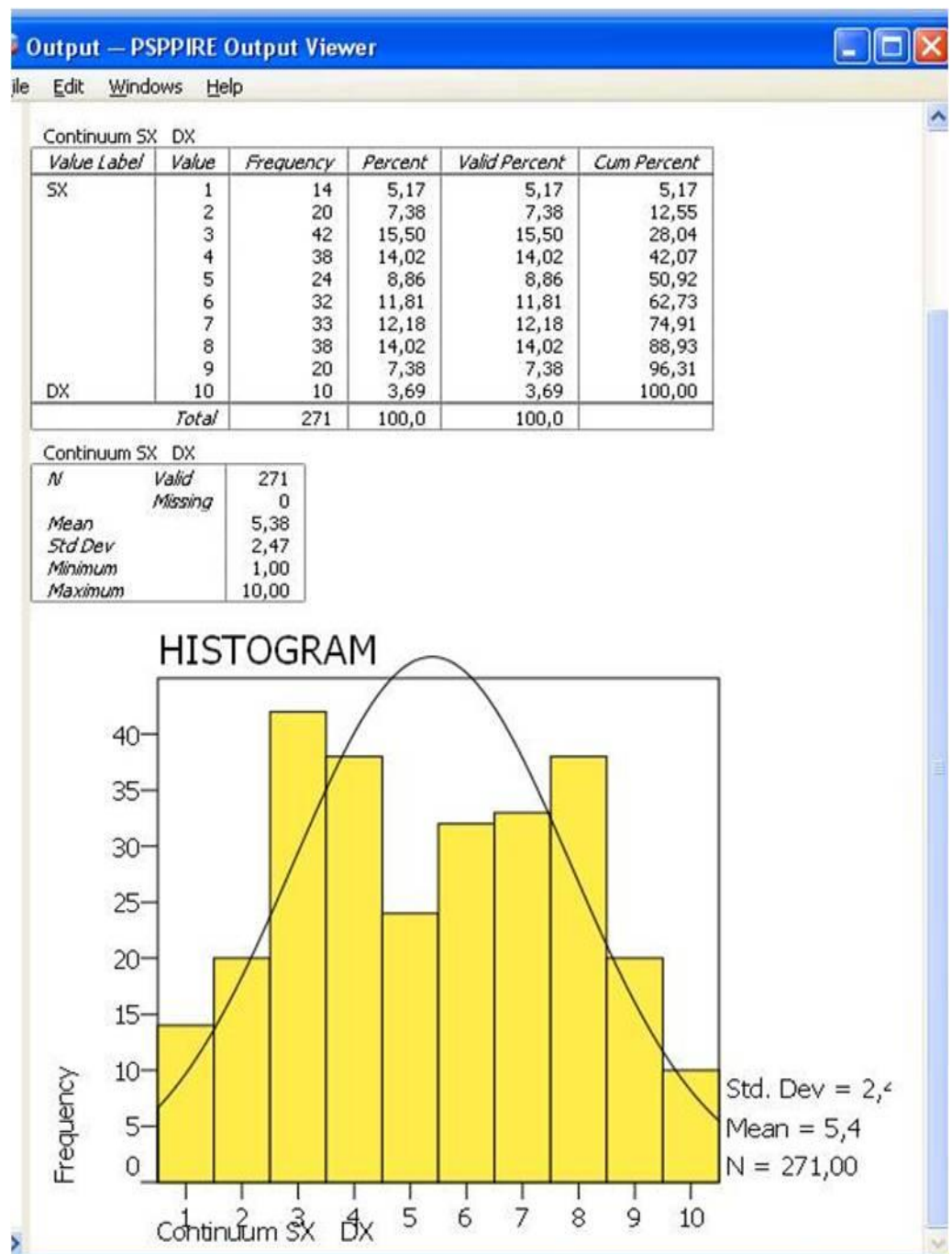
- cliccare su **Charts** e flaggare nella sua finestra le opzioni *Draw histograms* e *Superimpose normal curve*



- cliccare quindi sul bottone **OK** per ottenere il risultato voluto.

### Esito:

La finestra di Output mostra, nell'ordine:



- distrib. di frequenza

- Media, Scarto Tipo, Minimo e Massimo

- istogramma, con la approssimazione della curva normale



Da quanto sperimentato nei paragrafi precedenti, appare evidente come attraverso le procedure statistiche monovariate ***Frequencies*** e ***Descriptives*** sia possibile ottenere il massimo di informazione possibile su ogni variabile, sia essa una classificazione in senso stretto o un qualsiasi tipo di scala.

L'utente può quindi sperimentare a piacimento le due procedure, tenendo comunque conto della natura qualitativa o quantitativa delle variabili contenute nella matrice dati.

### C.5. Considerazioni propedeutiche agli ESERCIZI GUIDATI di analisi bivariata.

Nei prossimi paragrafi saranno proposti esercizi guidati riguardanti lo studio della relazione tra coppie di variabili presenti nella matrice dati. Occorre in proposito rinfrescarsi un attimo le idee sul significato di tali relazioni e sugli accorgimenti da rispettare, al fine di evitare di compiere operazioni prive di senso.

Va ricordato, in generale, come lo studio congiunto di due variabili NON necessiti di alcuna conoscenza statistica o matematica quando le stesse siano riferite ad un'unica unità osservativa. Analogamente a quanto visto per l'analisi monovariata, il singolo caso è infatti un'entità descrivibile attraverso una o molteplici modalità da lui espresse in riferimento a una o più proprietà.

Se, ad esempio, di un individuo di genere maschile si conoscono gli anni di età ( 39 ), il titolo di studio ( Master in economia ), la professione ( Consulente finanziario ), il partito per cui ha votato alle ultime elezioni ( Scelta Civica ) e quello per cui voterebbe domani ( PD ), si potrebbe dire che trattasi di una persona ancor giovane, di istruzione altamente qualificata, professionalmente attiva nel settore economico finanziario, che ha maturato uno spostamento del suo orientamento politico dal centro-moderato verso il centro-sinistra.

Aggiungendo l'osservazione delle opinioni dello stesso soggetto su questioni politiche e socio-culturali, si potrebbe arrivare a costruirne un "profilo individuale completo": il tutto, semplicemente osservando e interpretando il ventaglio di caratteristiche espresse.

Il problema nasce, viceversa, quando le proprietà da considerare singolarmente o incrociate fra loro fanno riferimento a  $n$  unità osservative, com'è di regola in una qualsiasi matrice dati rettangolare, casi ( righe ) per variabili ( colonne ).

Ora, se nell'analisi monovariata il problema è risolto contando i casi all'interno di ogni categoria di una variabile qualitativa o calcolando i valori medi per le variabili quantitative, come studiare la relazione congiunta [ bivariata ] tra due variabili?

La risposta è: utilizzando specifiche tecniche statistiche, ideate a seconda della natura qualitativa o quantitativa delle singole variabili costituenti la coppia.

A tal proposito, di ognuna se ne potrà banalmente riconoscere il tipo rispondendo alla domanda:

*ha senso calcolare la Media su questa variabile?*

Se la risposta è **Sì**, la variabile è di natura **quantitativa** ( genericamente: **scala** )

Se la risposta è **No**, la variabile è di natura **qualitativa** ( genericamente: **classificazione** )

NB. Fra le variabili su cui è possibile calcolare la Media vanno considerate, oltre alle scale in senso proprio ( come quelle enumerabili o frutto di una misurazione ), tutte le pseudo – scale originate da tecniche di *scaling* ( auto – ancoranti, termometro dei sentimenti, differenziali semantici, ecc. ), comprese le scale Likert. Di queste ultime, in particolare, occorre ricordare la duplice natura: qualitativa, dovuta alla relativa autonomia semantica delle categorie di risposta, basate su termini del linguaggio comune ( *contrario, poco d'accordo ... d'accordo* ); pseudo – quantitativa, conseguente all'assunzione di valore numerico attribuita ai codici corrispondenti alle categorie di risposta, ordinati da un minimo ad un massimo. Le scale Likert potranno dunque essere considerate indifferentemente scale o classificazioni.

A questo punto, le possibili combinazioni bivariante sono tre:

*Scala vs Scala*

*Scala vs Classificazione*

*Classificazione vs Classificazione*

Nel riquadro sottostante sono illustrate le tecniche statistiche da utilizzare in relazione a ciascun tipo di combinazione bivariata, corredata della conseguente procedura statistica richiamabile con PSPP [ vedi anche l'elenco del Menu **Analyze** , illustrato nel paragrafo **A.4.1.** ].

Tipo di relazione bivariata	Analisi statistica	Procedura PSPP
<i>Scala vs Scala</i>	Calcolo del coefficiente di correlazione.	<b><i>Bivariate <u>C</u>orrelation</i></b>
<i>Scala vs Classificazione</i>	Analisi della varianza. Calcolo di $\eta^2$ .	<b><i>Compare <u>M</u>eans</i></b>
<i>Classificazione vs Classificazione</i>	Tavola di contingenza, con frequenze attese, % sul totale o per <i>riga</i> o <i>colonna</i> . Eventuale $\Phi^2$ o Tau o altri per dicotomie.	<b><i><u>C</u>rosstabs</i></b>

### C.5.1. ESERCIZIO n° 4 : analisi bivariata tra due scale [ *Bivariate Correlation* ].

Si prendano in considerazione le variabili D10 [ *Continuum Sinistra Destra* ] e D11 [ *Frequenza alla messa* ]. La D11, come visibile dal questionario o dalle *Value Labels* nel registro delle variabili, è una classificazione con sei categorie ordinate secondo un criterio graduale di frequentazione del rito religioso della messa. Si va dal “mai” al “due o più volte la settimana”. Osservandola preliminarmente come classificazione ( **Analyze** ( → ) **Descriptive Statistics** ( → ) **Frequencies** ) si riscontra facilmente un basso livello di ritualismo.

La distribuzione di frequenza indica infatti che quasi la metà degli intervistati non va mai a messa e solo circa il 13% va a messa tutte le domeniche o più volte la settimana.

Questo è dunque il dato di partenza.

Frequenza alla messa					
Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
mai	1	128	47,23	47,23	47,23
2 - 3 v anno	2	62	22,88	22,88	70,11
1 v mese	3	16	5,90	5,90	76,01
2 - v mese	4	31	11,44	11,44	87,45
domeniche	5	31	11,44	11,44	98,89
2 o+ v sett	6	3	1,11	1,11	100,00
Total		271	100,0	100,0	

Ora, nell’esercitazione proposta, la variabile D11 sarà considerata come una pseudo – scala, avente estensione numerica da un minimo di 1 ad un massimo di 6 .

Al di là del generale basso livello di ritualismo, studiare la relazione di questa dimensione con l’autocollocazione politica significherà controllare empiricamente l’eventualità che possa esservi tra esse un legame interpretabile. L’assenza o presenza di tale legame sarà rivelato dal calcolo dell’indice di Correlazione di Pearson:  $r$  .

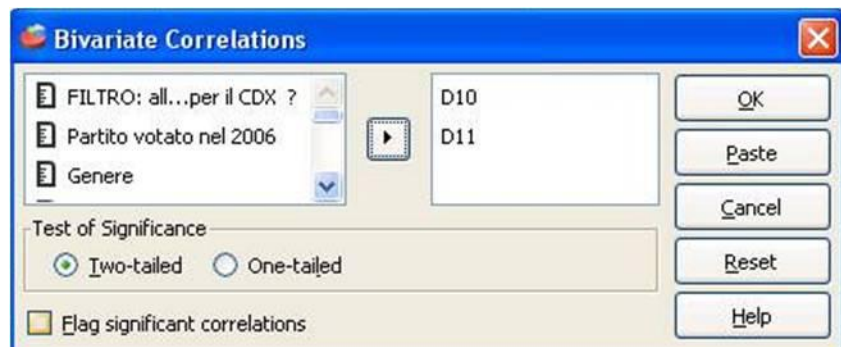
Come noto,  $r$  è un coefficiente *simmetrico*, indicante la strettezza della relazione tra due scale indipendentemente dall’influenza dell’una sull’altra. Assume valore  $\pm 1$  quando essa è massima, valore 0 quando la relazione è dovuta totalmente al caso. Il segno  $\pm$  sta ad indicarne la direzione: *direttamente proporzionale* ( + ), cioè al crescere di una variabile cresce anche l’altra; *inversamente proporzionale* ( - ), cioè al crescere di una variabile l’altra decresce.

Porre la richiesta a PSPP come segue:

- selezionare **Analyze** ( → ) **Bivariate Correlation** ( Si apre la relativa sotto-finestra );

- evidenziare una alla volta le etichette di D10 e D11 e spostare le variabili ( → ) nella *location* di destra;

- cliccare quindi sul bottone **OK** per ottenere il calcolo dell’  $r$  di Pearson .



L’esito della procedura è riassunta in una tabellina in cui compaiono le variabili correlate tra loro e, per effetto dell’automatismo del programma, pure con se stesse. Nella diagonale discendente compaiono infatti due  $r = 1$  in corrispondenza della correlazione tra ciascuna variabile e se stessa, mentre l’ $r$  calcolato tra le due variabili differenti è ripetuto nelle celle della diagonale ascendente.

**Notazione tecnica.** L’automatismo del programma riproducendo una tavola con valori ridondanti è dovuta al fatto che si potrebbero richiedere, con la stessa procedura statistica, le correlazioni bivariata fra più variabili. Nel prossimo esercizio guidato, in cui saranno richieste le correlazioni fra le frequenze di ascolto dei sette telegiornali, si avrà modo di apprezzare l’efficacia sintetica dell’Output automatico.

**Commento.** Il valore di  $r$ , pari a  $+0,23$ , rivela un nesso tra le due variabili, indicante una maggiore frequentazione del rito religioso in corrispondenza di un posizionamento politico individuale orientato verso destra e, viceversa, una frequenza nulla o ridotta in coincidenza di posizionamenti politici orientati a sinistra. Tale legame, letto in chiave socio-politica, può essere interpretato come indizio della mancata piena secolarizzazione del nostro elettorato.

l'Italia appare dunque un po' più distante dai paesi di consolidata tradizione liberal democratica e più vicina, con i dovuti distinguo, ai paesi in cui la religione gioca un ruolo attivo nell'indirizzo delle scelte della classe politica, sia di governo che di opposizione.

Output – PSPPIRE Output Viewer

File Edit Windows Help

CORRELATIONS

CORRELATION

/VARIABLES = D10 D11

/PRINT = TWOTAIL SIG.

Correlations

		Continuum SX DX	Frequenza alla messa
Continuum SX DX	Pearson Correlation	1,00	,23
	Sig. (2-tailed)		,000
	N	271	271
Frequenza alla messa	Pearson Correlation	,23	1,00
	Sig. (2-tailed)	,000	
	N	271	271

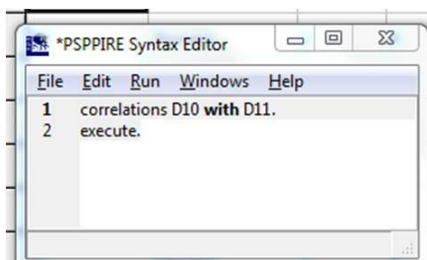
Quanto argomentato sollecita un approfondimento del legame tra autocollocazione politica e ritualismo. Se, come indicato dal valore di  $r$ , esiste tra loro un nesso debole ma comunque significativo, occorre riquilibrare il modello di relazione come *asimmetrico*. Ciò significa considerare il comportamento ritualistico come variabile *indipendente* e l'autocollocazione politica come variabile da esso *dipendente*.

A livello statistico, attraverso il Coefficiente di determinazione  $R^2$  si potrà misurare la frazione di devianza spiegata, cioè la proporzione di variabilità dell'*Autocollocazione politica* "spiegata" dalla *Frequenza alla messa*. Sarà dunque sufficiente moltiplicare il valore di  $r$  per se stesso ( $0,23 \times 0,23 = 0,0529$ ) e concludere che in poco più del 5% dei soggetti intervistati la decisione di auto collocarsi sul *Continuum* è associabile al loro comportamento ritualistico. E' tanto o poco? Né l'una né l'altra cosa: è infatti plausibile che tale associazione sia comunque ridotta, a fronte di una società ampiamente secolarizzata, così come, quando essa si manifesti, sia il sintomo di tratti culturali profondi resistenti al trascorrere del tempo.

Tornando alla questione della ridondanza dell'automatismo riprodotto nell'output, va detto che può essere evitata ricorrendo all'uso dei comandi nella finestra di Sintassi. Sarà sufficiente scrivere:

CORRELATION D10 WITH D11

per ottenere la tabellina riassuntiva riportante solamente il valore di  $r$  calcolato tra le due variabili.



Run

(→)

Selection

Correlations

		Frequenza alla messa
Continuum SX DX	Pearson Correlation	,23
	Sig. (2-tailed)	,000
	N	271

### C.5.2. ESERCIZIO n° 5 : analisi bivariata tra più scale [ *Bivariate Correlation* ].

Calcolare le correlazioni bivariate fra gli ascolti di ciascuno dei sette telegiornali. Indifferentemente dall'utilizzo del Menu **Analyze** (→) **Bivariate Correlation** o del comando **CORRELATIONS** fatto eseguire in una finestra di sintassi, PSPP produrrà una tabella quadrata di  $7 \times 7 = 49$  celle, con 7 correlazioni di valore 1 (ogni variabile con se stessa) nella diagonale discendente e 21 valori differenti, ripetuti al di sotto e al di sopra della diagonale, relativi alle correlazioni bivariate tra un TG e i rimanenti sei.

		TG1, Frequenza ascolto	TG2, Frequenza ascolto	TG3, Frequenza ascolto	TG4, Frequenza ascolto	TG5, Frequenza ascolto	Studio Aperto, Frequenza ascolto	TG la 7, Frequenza ascolto
TG1, Frequenza ascolto	Pearson Correlation Sig. (2-tailed) N	1,00 271	,50 ,000 271	,37 ,000 271	,03 ,678 271	-,04 ,530 271	-,09 ,128 271	,06 ,366 271
TG2, Frequenza ascolto	Pearson Correlation Sig. (2-tailed) N	,50 ,000 271	1,00 ,000 271	,39 ,000 271	,11 ,084 271	,06 ,328 271	,08 ,198 271	,06 ,341 271
TG3, Frequenza ascolto	Pearson Correlation Sig. (2-tailed) N	,37 ,000 271	,39 ,000 271	1,00 ,000 271	-,09 ,133 271	-,16 ,009 271	-,17 ,004 271	,31 ,000 271
TG4, Frequenza ascolto	Pearson Correlation Sig. (2-tailed) N	,03 ,678 271	,11 ,084 271	-,09 ,133 271	1,00 ,000 271	,41 ,000 271	,32 ,000 271	,09 ,128 271
TG5, Frequenza ascolto	Pearson Correlation Sig. (2-tailed) N	-,04 ,530 271	,06 ,328 271	-,16 ,009 271	,41 ,000 271	1,00 ,000 271	,42 ,000 271	-,07 ,231 271
Studio Aperto, Frequenza ascolto	Pearson Correlation Sig. (2-tailed) N	-,09 ,128 271	,08 ,198 271	-,17 ,004 271	,32 ,000 271	,42 ,000 271	1,00 ,000 271	,03 ,662 271
TG la 7, Frequenza ascolto	Pearson Correlation Sig. (2-tailed) N	,06 ,366 271	,06 ,341 271	,31 ,000 271	,09 ,128 271	-,07 ,231 271	,03 ,662 271	1,00 271

**Commento.** Dei ventuno coefficienti prodotti dalla procedura (osservabili per comodità al di sotto della diagonale) solo sette hanno valori di  $r$  abbastanza elevati. Sono infatti positivamente e reciprocamente correlati gli ascolti dei telegiornali delle tre reti nazionali, così come i tre TG della galassia Mediaset. Il TG La7, emittente indipendente, si correla significativamente solo con il TG3. Questi valori delineano potenziali aree omogenee di ascolto anche in relazione al tipo di orientamento politico degli intervistati.

#### C.5.2.1. ESERCIZIO n° 6 : analisi bivariata per controllo ipotesi.

L'ipotesi sopra suggerita può essere controllata calcolando l'indice di correlazione tra le frequenze d'ascolto ai singoli TG e l'autocollocazione sul *Continuum Sinistra Destra* [ variabile D10 ].

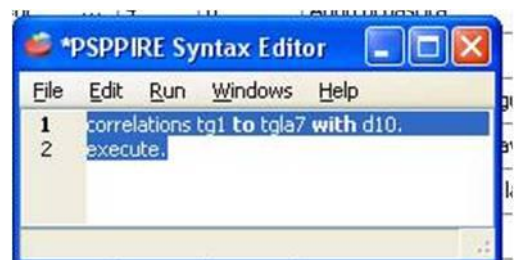
In questo caso è opportuno ricorrere alla Sintassi.

Procedere pertanto così:

- cliccare su

**File** (→) **New** (→) **Syntax** e

- scrivere il comando **CORRELATIONS**, specificando la lista dei TG e la *keyword* **WITH** seguita dalla variabile D10.



**Esito:**

I sette valori di  $r$  calcolati tra la frequenza d'ascolto dei singoli TG e l'auto posizionamento sul *Continuum Sinistra → Destra* indicano coerentemente valori elevati e di segno positivo tra i servizi dei canali Mediaset e l'autocollocazione sulla destra del Continuum. Come cartina di tornasole, un valore di  $r$  elevato ma di segno negativo con il TG3. TG1, TG2 e La7 hanno invece valori pressoché nulli, pur con segni negativi. Il risultato può dunque essere letto come indizio di "ascolti militanti" da parte dei tifosi di Silvio Berlusconi da una parte e dei soggetti della sinistra più tradizionale dall'altra. Il resto dei TG appare invece più coerente con la *mission* di informazione rivolta a tutte le platee.

Output — PSPPIRE Output Viewer

File Edit Windows Help

CORF CORRELATIONS  
EXEC correlations tg1 to tgl7 with d10.

Correlations

		Continuum SX DX
TG1, Frequenza ascolto	Pearson Correlation	,15
	Sig. (2-tailed)	,012
	N	271
TG2, Frequenza ascolto	Pearson Correlation	-,06
	Sig. (2-tailed)	,319
	N	271
TG3, Frequenza ascolto	Pearson Correlation	-,40
	Sig. (2-tailed)	,000
	N	271
TG4, Frequenza ascolto	Pearson Correlation	,51
	Sig. (2-tailed)	,000
	N	271
TG5, Frequenza ascolto	Pearson Correlation	,56
	Sig. (2-tailed)	,000
	N	271
Studio Aperto, Frequenza ascolto	Pearson Correlation	,42
	Sig. (2-tailed)	,000
	N	271
TG la 7, Frequenza ascolto	Pearson Correlation	-,07
	Sig. (2-tailed)	,270
	N	271

EXECUTE  
execute.

### C.5.3. ESERCIZIO n° 7 : analisi bivariata tra scala e classificazione [ *Compare Means* ].

Studiare la relazione tra D11 [ *Frequenza alla messa* ] e AggD2 [ comportamento di voto nel 2006 ]. L'*ipotesi nulla* sottesa a questo modello di relazione sarà confermata qualora il valore della Media del comportamento ritualista calcolato nei singoli gruppi di elettori che hanno votato differenti partiti o non hanno votato sarà all'incirca identico al valore della Media calcolato su tutti i 271 casi. La conferma di tale ipotesi sancirebbe la casualità della relazione tra le due variabili. Al contrario, valori medi di categoria differenti sarebbero l'indizio di un legame interpretabile. Il primo passo è allora quello di valutare l'entità degli scostamenti delle medie di categoria da quella generale.

Procedere quindi come segue. **Preliminarmente:**

- trovarsi nella modalità visualizzazione variabili;
- posizionarsi sulla riga corrispondente ad AggD2 e
- dare un clic sui tre puntini ... in corrispondenza della colonna *Missing Values*;
- flaggare l'opzione **No missing values** e
- dare quindi l'OK .

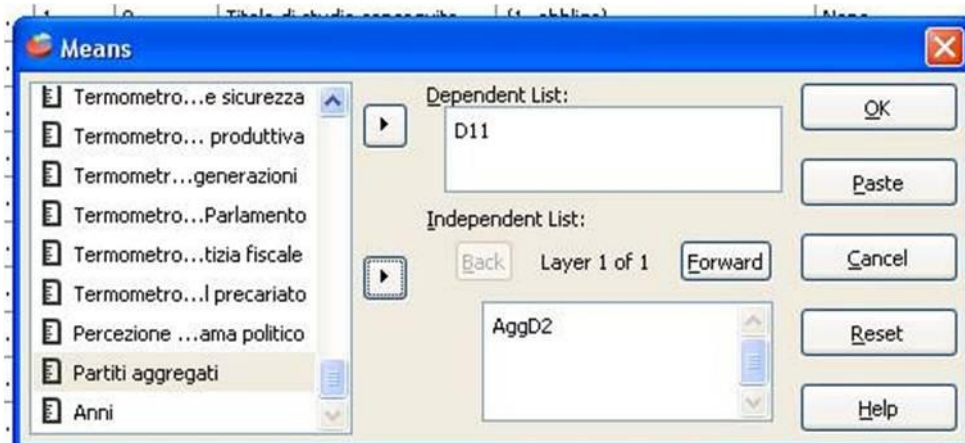
La AggD2 sarà ora disponibile **anche** per i 51 soggetti intervistati che hanno dichiarato di non aver votato nel 2006.



Ora è possibile richiedere l'incrocio tra la scala D11 e la classificazione AggD2. Prima di tutto richiamare la procedura:

- cliccare su **Analyze** ( → ) **Compare Means** ( → ) **Means**

**Esito:**



Dopodiché:

- evidenziare l'etichetta *Frequenza alla messa* e spostare D11 ( → ) nella Dependent List;
- evidenziare l'etichetta *Partiti aggregati* e spostare AggD2 ( → ) nella Independent List ;
- dare quindi l'OK .



L'output fornisce i valori di Media e Scarto Tipo della D11 per tutti i gruppi di soggetti che hanno votato lo stesso partito e per i 51 astenutisi. I valori delle singole medie, oltre ad essere comparati fra loro, dovranno essere confrontati con la Media pari a 2,20 calcolata sull'intera popolazione. Questo valore è piuttosto basso, poiché i "gradini" 2 e 3 della D11 corrispondono alle modalità *Due – tre volte l'anno* e *Una volta al mese*. Come peraltro già commentato all'inizio del paragrafo C.5.1., il ritualismo religioso indicato dall'assiduità nell'andare a messa appare dunque abbastanza circoscritto.

MEANS  
MEANS TABLES = D11  
BY AggD2.

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Frequenza alla messa: Partiti aggregati	271	100%	0	0%	271	100%
Frequenza alla messa:	271	100%	0	0%	271	100%

Report

	Partiti aggregati	Mean	N	Std. Deviation
Frequenza alla messa	DS	1,72	60,00	1,12
	MARGHERITA	2,00	14,00	1,52
	PRC	1,83	18,00	1,29
	ALTRO CSX	1,61	18,00	1,29
	FI	2,80	44,00	1,56
	AN	2,37	41,00	1,34
	LEGA	2,30	10,00	1,49
	UDEUR	3,80	15,00	1,57
	Non voto	2,04	51,00	1,50

Report

	Mean	N	Std. Deviation
Frequenza alla messa	2,20	271,00	1,47

**Commento.** La media della D11 segna il suo apice nel gruppo di 15 soggetti che hanno votato UDEUR (scarto di +1,56 sul valore medio totale) e in quello numericamente più consistente dei forzisti (+0,56). Sul fronte opposto, i valori più bassi sono calcolati nei gruppi della sinistra radicale (-0,63) e dei DS (-0,52). Da notare, in ogni caso, come i partiti dell'area di CDX siano tutti al di sopra della media totale, mentre quelli di CSX, insieme al gruppo degli astenuti, siano tutti al di sotto. Interessante anche la media di 2,0 degli elettori della MARGHERITA, testimoniante il processo di secolarizzazione interna all'area politica della ex sinistra democristiana.

### C.5.4. ESERCIZIO n° 8 : analisi bivariata tra scala e classificazione [ *One Way ANOVA* ].

Quanto osservato nel paragrafo precedente ( differenza fra le medie della *Frequenza alla messa* calcolate sui casi articolati secondo il *Partito votato nel 2006* ) ha permesso di argomentare sul nesso esistente tra il – comunque relativamente basso - ritualismo cattolico e l'orientamento politico.

A livello di analisi statistica, il fenomeno sopra descritto può essere ulteriormente caratterizzato mediante il calcolo del coefficiente sintetico  $Eta^2$ , ricavabile dalla procedura denominata Analisi della Varianza Esterna [ ANOVA ].

Chi desideri ottenere, oltre alla differenza tra le medie, anche il contributo dell'ANOVA, può evitare quindi la procedura *Means* vista nel paragrafo precedente, richiedendo direttamente a PSPP la *Compare Means* nella sua variante *One Way ANOVA* .

Questi i passi da seguire. Cliccare su:

- *Analyze* ( → ) *Compare Means* e selezionare nel sotto menu

- *One Way ANOVA* quindi:

- evidenziare l'etichetta della D11

- dare un clic su ( → ) per spostare D11 in *Dependent Variable(s)*;

- evidenziare l'etichetta di AggD2

- dare un clic su ( → ) per spostare AggD2 nel campo *Factor*

- flaggare *Descriptives* nell'area *Statistics*

- dare infine l'OK .



**Esito:**

Output – PSPPIRE Output Viewer										
File Edit Windows Help										
ONEWAY										
ONEWAY /VARIABLES= D11 BY AggD2										
/STATISTICS=DESCRIPTIVES .										
Descriptives										
		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum	
						Lower Bound	Upper Bound			
<i>Frequenza alla messa</i>	DS	60	1,72	1,12	,14	1,43	2,01	1	5	
	MARGHERITA	14	2,00	1,52	,41	1,12	2,88	1	5	
	PRC	18	1,83	1,29	,31	1,19	2,48	1	5	
	ALTR0 CSX	18	1,61	1,29	,30	,97	2,25	1	5	
	FI	44	2,80	1,56	,24	2,32	3,27	1	6	
	AN	41	2,37	1,34	,21	1,94	2,79	1	5	
	LEGA	10	2,30	1,49	,47	1,23	3,37	1	5	
	UIDEUR	15	3,80	1,57	,40	2,93	4,67	1	6	
	Non voto	51	2,04	1,50	,21	1,62	2,46	1	5	
	Total	271	2,20	1,47	,09	2,03	2,38	1	6	
ANOVA										
		Sum of Squares	df	Mean Square	F	Sig.				
<i>Frequenza alla messa</i>	Between Groups	79,78	8	9,97	5,18	,000				
	Within Groups	504,05	262	1,92						
	Total	583,84	270							

Oltre alle medie di categoria ricavabili dalla semplice **Means**, l'ANOVA calcola:

- la devianza totale ( o TSS, dall'inglese *Total Sum of Squares* ), cioè la sommatoria dei quadrati degli scarti ottenuti sottraendo dal valore di ogni caso quello della media della popolazione;

- la devianza *interna* ( o WSS, *Within Sum of Squares* ), cioè la sommatoria dei quadrati degli scarti tra ogni singolo valore reale e la media della categoria di appartenenza;

- la devianza *esterna* ( o BSS, *Between Sum of Squares* ), cioè la sommatoria dei quadrati – moltiplicati per il numero di casi presenti in ogni categoria – degli scarti ottenuti sottraendo dal valore di ciascuna media di categoria quello relativo alla media della popolazione.

La devianza totale della scala è dunque esprimibile dalla somma di due devianze, quella interna alle categorie e quella tra le categorie e la media generale; quindi:  $TSS=WSS+BSS$  (  $583,84=504,05+79,78$  ).

Con questi elementi a disposizione, diventa semplice calcolare il valore di  $Eta^2$ , espresso dalla formula

$$Eta^2 = \frac{BSS}{TSS}$$

**NB.** Tanto più grande è WSS tanto più piccolo diventa, conseguentemente, il contributo di BSS alla variabilità totale della scala. Infatti, prendendo spunto dai due casi limite, quando  $Eta^2 = 1$  ( cioè  $BSS = TSS$  ), WSS sarà uguale a 0; mentre quando  $Eta^2 = 0$  ( cioè  $BSS = 0$  ), WSS sarà uguale a TSS, indicando che tutta la variabilità della scala è **indipendente** dall'articolazione categoriale della classificazione (cioè tutte le categorie hanno la stessa media).

$Eta^2$  va interpretato come la quota di variabilità della scala attribuibile all'articolazione categoriale prodotta dalla classificazione. Considerando i valori di BSS ( 79,78 ) e TSS ( 504,05 ) si ottiene:

$$Eta^2 = \frac{79,78}{583,84} = 0,14$$

Il valore di 0,14 indica pertanto che il 14% del ritualismo è associabile al comportamento di voto tenuto dagli intervistati.

**Limite tecnico del programma.** La procedura **One Way ANOVA** fornisce i valori di BSS e TSS necessari al calcolo di  $Eta^2$  ma, nella corrente versione di PSPP per Windows, tale calcolo va fatto dall'utente. Viceversa, il padre nobile SPSS contempla in automatico anche il valore di  $Eta^2$ .

## C.6. Premesse allo studio congiunto di due classificazioni.

Lo studio congiunto ( o *incrocio* ) di coppie di classificazioni rappresenta un tipo di analisi bivariata molto frequente nelle ricerche socio - politologiche. Le questioni trattate in tali ambiti si riferiscono infatti prevalentemente a concetti complessi o a proprietà semplici di natura qualitativa, come tali rilevate. In alcuni casi, i ricercatori si svincolano dalla natura qualitativa delle proprietà, aggirando il limite mediante l'utilizzo di pseudo - scale originate da molteplici tecniche di *scaling* ( come visto a proposito delle scale *auto ancoranti*, del *termometro dei sentimenti* e delle *scale Likert* ). In tutti gli altri casi, le proprietà qualitative vengono invece rilevate e definite operativamente mediante classificazioni.

Diventa pertanto importante avere le idee chiare sui presupposti metodologici ( cosa accade nell'incrocio tra due classificazioni ) e sul tipo di varianti statistiche da scegliere, onde ottenere le informazioni per una migliore comprensione della relazione.

### C.6.1. Significato dell'incrocio tra due classificazioni.

E' innanzitutto utile ricordare la caratteristica saliente di una classificazione: la piena autonomia semantica delle modalità in cui essa si manifesta. Conseguentemente, ogni categoria ha in comune con le altre il solo fatto di essere uno *stato possibile* della classificazione. Questo è il motivo per cui, come già visto, l'analisi statistica monovariata si limita al conteggio dei casi all'interno di ciascuna categoria.

A livello teorico - date le premesse di cui sopra - l'incrocio di due classificazioni produrrà  $n$  modalità tipologiche, rappresentate da tutte le possibili coppie di modalità congiunte. Di ogni modalità tipologica si potranno allora conteggiare le frequenze ( numero dei casi ), provando quindi ad interpretarne la differente consistenza.

Nel concreto, l'incrocio sarà rappresentato da una tavola a doppia entrata ( o *tavola di contingenza* o *tabulazione incrociata* ), avente un numero di celle pari al prodotto del numero di modalità delle due classificazioni.

Se, ad es., una variabile articola i soggetti in base a cinque possibili condizioni rispetto al lavoro [ *studente, disoccupato, casalinga/o, in attività lavorativa, pensionato* ] e un'altra classifica gli stessi individui rispetto all'aver partecipato alle ultime elezioni politiche [ *ha votato, non ha votato* ], la relazione tra le due variabili andrà letta osservando come si distribuiscono gli  $N$  intervistati nelle (  $5 \times 2$  ) dieci modalità tipologiche : **studente che ha votato**; **studente che non ha votato** ... e così via, fino a ... **pensionato che ha votato**; **pensionato che non ha votato**.

		Condizione rispetto al lavoro					
Partecipazione al voto alle ultime elezioni		<i>studente</i>	<i>disoccupato</i>	<i>Casalinga/o</i>	<i>In attività</i>	<i>pensionato</i>	Totale
<i>Ha votato</i>							T <i>Ha votato</i>
<i>Non ha votato</i>							T <i>Non ha votato</i>
Totale		T <i>studenti</i>	T <i>disoccupati</i>	T <i>casalinghe/i</i>	T <i>in attività</i>	T <i>pensionati</i>	N

Gli individui si distribuiranno presumibilmente con una differente numerosità nelle dieci combinazioni. Si dovrà allora rispondere al quesito:

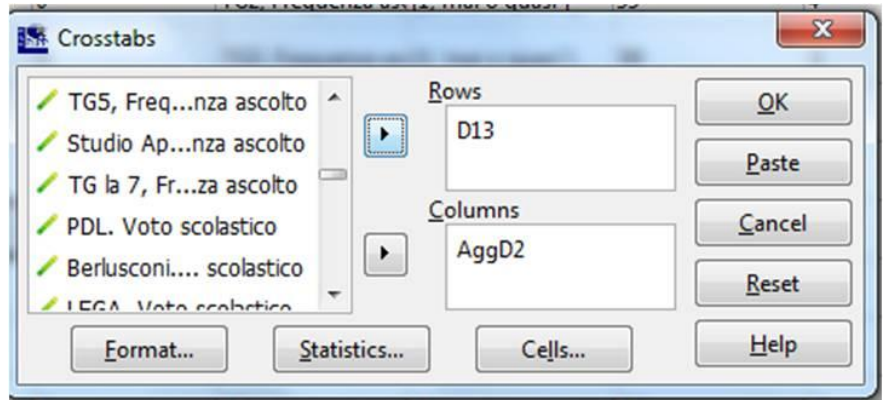
*Come fare a cogliere un nesso tra le due variabili osservando i conteggi all'interno di ciascuna cella ?*  
Come fare, in altri termini, a comprendere se il numero di soggetti per ciascuna modalità tipologica è dovuta al caso o indica invece legami interpretabili?

La risposta al quesito verrà trovata nel prossimo esercizio, che sarà svolto secondo tre criteri di esecuzione differenti.

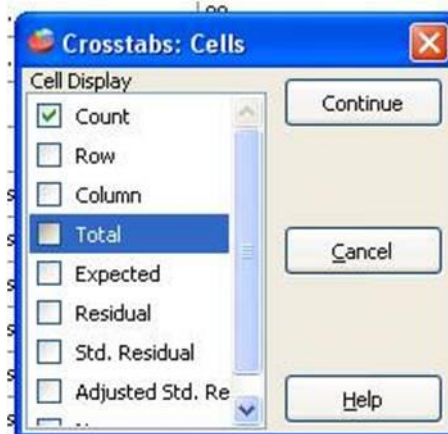
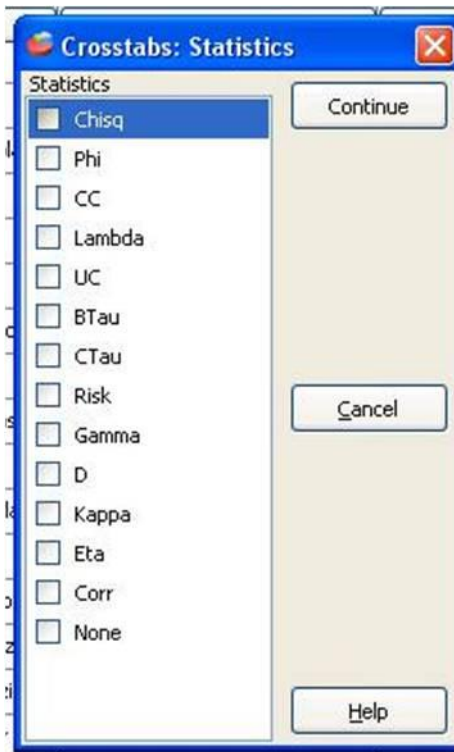
### C.7. ESERCIZIO n° 9 : incrocio grezzo tra due classificazioni [ *Crosstabs* ].

Riprodurre l'incrocio tra la variabile D13 ( *Comportamento alle elezioni 2008* ) e la AggD2 ( *Partiti aggregati*, votati nel 2006 ). La tavola di contingenza sarà composta da  $4 \times 9 = 36$  modalità tipologiche. Procedere come segue:

- Attivare la procedura **Analyze** ( → ) **Descriptive Statistics** ( → ) **Crosstabs** ;
- evidenziare l'etichetta *Comportamento alle elezioni 2008* e spostare ( → ) la D13 nel campo **Rows**;
- evidenziare l'etichetta *Partiti aggregati* e spostare ( → ) AggD2 nel campo **Columns**.



A questo punto occorre intervenire nei campi relativi alle funzioni **Statistics** e **Cells** . Quindi, nell'ordine:



- cliccare su **Statistics** ,
- deflaggare la voce **Chisq** e
- cliccare su **Continue**.
- Cliccare poi su **Cells** e
- deflaggare le voci **Row**, **Column** e **Total**, lasciando attiva la sola voce **Count**;
- cliccare su **Continue** per tornare alla finestra iniziale del comando.

Dare quindi l'**OK** per eseguire la procedura statistica.

L'esito è il seguente:

Comportamento alle elezioni 2008	Partiti aggregati									Total
	DS	MARGHERITA	PRC	ALTRO CSX	FI	AN	LEGA	UDEUR	Non voto	
sa chi voterà	39,00	6,00	3,00	6,00	26,00	25,00	6,00	5,00	5,00	121,00
area si, partito no	11,00	5,00	9,00	6,00	10,00	13,00	,00	5,00	8,00	67,00
indeciso	7,00	2,00	4,00	3,00	3,00	1,00	2,00	3,00	13,00	38,00
rifiuto	3,00	1,00	2,00	3,00	5,00	2,00	2,00	2,00	25,00	45,00
Total	60,00	14,00	18,00	18,00	44,00	41,00	10,00	15,00	51,00	271,00

**Osservazione.** La tavola di contingenza produce i **conteggi delle frequenze reali** di ogni modalità tipologica [ d'ora in avanti: **fr** ] all'interno di ogni combinazione ( cella ) prodotta dall'incrocio. Oltre a ciò, sono indicati i subtotali di ciascuna categoria, sia in riga ( D13 ) sia in colonna ( AggD2 ), così come il Totale generale, pari a 271 .

**Interpretazione dei conteggi.** La differente numerosità di casi riscontrabile nelle 36 celle, unitamente allo squilibrio fra i subtotali delle quattro categorie della variabile D13 e delle nove categorie della AggD2, consente di azzardare solamente interpretazioni plausibili e, in ogni caso, approssimative. Oltre a ciò, è bene sottolineare l'importanza di scegliere un criterio di lettura dei dati. Evitare una lettura a *casaccio* e scegliere l'orientamento *per riga* o *per colonna* può aiutare nel tentativo di interpretazione del contenuto della tavola di contingenza. La scelta dell'orientamento di lettura comporterà ovviamente valutazioni differenti, anche se riferite alle stesse **fr** .

**Letture per riga.** In questo caso, ogni singola **fr** va comparata con le altre **fr** giacenti nelle celle della stessa riga e con il subtotale di casi della categoria espressa nella riga stessa.

Ad es., la prima riga della tavola di contingenza indicherà come i 121 soggetti che *sanno quale partito votare alle successive elezioni del 2008* si distribuiscono secondo il *partito votato nel 2006*. Va da sé che le **fr** risentiranno della differente consistenza numerica dei partiti votati nel 2006 ( i cui subtotali si trovano in colonna ). Nel caso specifico, si riscontra come i tre partiti col numero di frequenze maggiori riproducano l'ordine di grandezza dei loro totali marginali ( 39 DS, 26 FI e 25 AN ) mentre gli altri violano parzialmente questa corrispondenza.

La successiva osservazione *per riga* delle **fr** riferite alle altre categorie *area sì – partito no, indeciso e rifiuto* diventa infine più azzardata a causa della loro minore consistenza numerica.

In ogni caso, il tipo di lettura per riga risponde alla domanda:

*Rispetto alla scelta di ogni specifica modalità del comportamento di voto per il 2008, come si distribuiscono i soggetti in base al partito votato nel 2006 ?*

**Letture per colonna.** In questo caso, ogni singola **fr** va comparata con le altre **fr** giacenti nelle celle della stessa colonna e con il subtotale di casi della categoria espressa nella colonna.

Ad es., osservando la colonna dei soggetti che avevano votato DS nel 2006, si potrà affermare che 39 su 60 hanno dichiarato la conferma del voto nelle imminenti elezioni, 14 (11+3) cambieranno partito, mentre 7 sono indecisi. Ripetendo l'osservazione sulle altre colonne si potrà congetturare quali partiti guadagneranno o perderanno. Analogamente a quanto visto prima, le interpretazioni sui partiti minori comportano un'ulteriore assunzione di rischio a causa della loro scarsa consistenza numerica.

In ogni caso, il tipo di lettura per colonna risponde alla domanda:

*I soggetti distinti secondo il partito votato nel 2006 come si distribuiscono in base alla scelta delle modalità di comportamento di voto per il 2008 ?*

Alla fine, indipendentemente dal tipo di lettura scelto, e assumendo che i subtotali delle categorie delle due variabili siano differenti, rimangono incontestabili due fatti:

- le **fr** prodotte nella tavola di contingenza **rappresentano la REALTA' della relazione** tra le due variabili;
- limitando l'osservazione alle sole **fr** , viene a mancare un criterio di comparazione, in grado di svelare se tali frequenze indicano o meno squilibri/tendenze all'interno di una o più celle; indizi di una relazione interpretabile tra le due classificazioni.

Nei prossimi due esercizi saranno testate altrettante varianti statistiche della procedura **Crosstabs** , in grado di rispondere alla necessità di disporre di un efficace criterio di comparazione.

### C.8. ESERCIZIO n° 10. Ipotesi nulla: Crosstabs con frequenze attese [ *Expected* ].

**Premessa.** L'esercizio consisterà nell'esecuzione della procedura Crosstabs, corredata dell'opportuna variante [ *Expected* ]. Tale variante affianca a ciascuna **fr** una **frequenza teorica attesa** [ d'ora in avanti: **ft** ] ottenuta in modo tale da rispettare la consistenza dei subtotali delle categorie riferite alle due variabili, normalizzate per il numero totale dei casi.

Le **ft**, nella sostanza, riproducono virtualmente la situazione soggiacente all'ipotesi nulla, in base alla quale la relazione tra le due variabili è legata solamente al loro modo congiunto ma indipendente di manifestarsi.

Ogni **ft** verrà calcolata moltiplicando il corrispondente subtotale di riga per il subtotale di colonna e dividendo il prodotto per il numero Totale dei casi.

Stando all'esempio concreto delle variabili D13 e AggD2, i dati di partenza per calcolare le **ft** sono pertanto i reali subtotali di **riga** e **colonna** e il **Totale dei casi**, come qui sotto raffigurato.

	DS	MAR	PRC	AI-CSX	FI	AN	LEGA	UDEUR	Non Voto	n
Sa chi voterà										121
Area sì, partito no										67
Indeciso			fr 4 ft 3		fr 3 ft 6					38
Rifiuto										45
n	60	14	18	18	44	41	10	15	51	N=271

Applicando l'algoritmo alle modalità tipologiche *PRC + indeciso* e *FI + indeciso* otterremo:

$$18 \times 38 / 271 = 2,5239... * \quad \text{e} \quad 44 \times 38 / 271 = 6,1697... *$$

\* I quozienti sono, per definizione, numeri puri: presentano un valore intero seguito da *n* decimali. Bisogna tuttavia considerare che le frequenze attese approssimano CASI, cioè oggetti indivisibili. Occorrerà pertanto **arrotondare** la **ft**, così da stabilire il totale di casi teorici attesi nella singola cella.

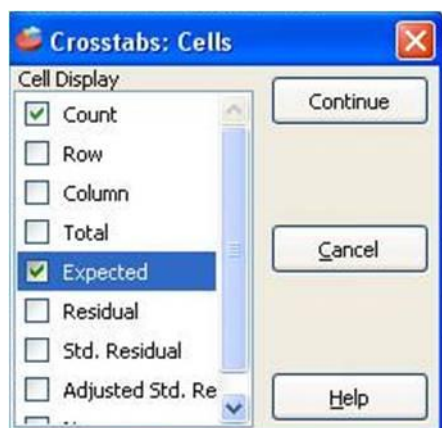
Se l'ipotesi nulla fosse confermata, dovremmo quindi avere: tre soggetti indecisi, fra i 18 intervistati che nel 2006 avevano votato PRC; sei soggetti indecisi, fra i 44 intervistati che nel 2006 avevano votato FI.

In questo caso, l'ipotesi nulla viene smentita dalla realtà. Confrontando i dati reali con quelli teorici, si può infatti asserire che gli indecisi di provenienza PRC sono più di quanti avrebbero dovuto esserci in presenza di una relazione casuale tra le due variabili, mentre gli indecisi di area forzista sono la metà di quelli attesi. Gli scarti **fr - ft** ( PRC = **+1**; FI = **-3** ) sarebbero dunque indizio di una incrinatura della fedeltà dell'elettorato PRC e, viceversa, di una forte coesione in FI. Ciò deve ovviamente essere confermato dall'osservazione di altri scarti.

Va infine ribadito come la validità dell'ipotesi nulla ( cioè della relazione casuale verificabile nelle modalità tipologiche ) sarà confermata ogniqualevolta gli scarti **fr - ft** saranno uguali a zero.

**ESERCIZIO.** Eseguire i seguenti passi:

- Attivare la procedura Analyze ( → ) Descriptive Statistics ( → ) Crosstabs e, come fatto prima,
- evidenziare le etichette e spostare ( → ) la D13 e la AggD2 nei campi Rows e Columns.
- passare quindi, in successione, nei campi delle funzioni Statistics e Cells ;



- nel campo **Statistics** deflaggare **Chsq** ;
- passare alla funzione **Cells** ( qui illustrata a sinistra );
- deflaggare tutte le opzioni tranne la **Count** e
- flaggare l'opzione **Expected** ;
- cliccare su *Continue* e tornare alla finestra iniziale del comando.
- Cliccare sul bottone **OK** per eseguire la procedura statistica.

PSPP riproduce la stessa tavola di contingenza vista nell'esercizio precedente con l'aggiunta, al di sotto del valore reale [ *count* ], del valore teorico atteso [ *expected* ].

Comportamento alle elezioni 2008	Partiti aggregati									Total
	DS	MARGHERITA	PRC	ALTRO CSX	FI	AN	LEGA	UDEUR	Non voto	
sa chi voterà	39,00	6,00	3,00	6,00	26,00	25,00	6,00	5,00	5,00	121,00
area sì, partito no	26,79	6,25	8,04	8,04	19,65	18,31	4,46	6,70	22,77	,00
indeciso	11,00	5,00	9,00	6,00	10,00	13,00	,00	5,00	8,00	67,00
rifiuto	14,83	3,46	4,45	4,45	10,88	10,14	2,47	3,71	12,61	,00
	7,00	2,00	4,00	3,00	3,00	1,00	2,00	3,00	13,00	38,00
	8,41	1,96	2,52	2,52	6,17	5,75	1,40	2,10	7,15	,00
	3,00	1,00	2,00	3,00	5,00	2,00	2,00	2,00	25,00	45,00
	9,96	2,32	2,99	2,99	7,31	6,81	1,66	2,49	8,47	,00
Total	60,00	14,00	18,00	18,00	44,00	41,00	10,00	15,00	51,00	271,00

Con i dati disponibili è quindi possibile calcolare manualmente la differenza **fr - ft** in tutte le 36 celle e, sulla base dell'entità degli scarti positivi o negativi di ciascuna – in questo caso scegliendo di leggerli *per colonna* - operare una stima dei guadagni e delle perdite prevedibili per ciascun partito di riferimento.

	DS	MAR	PRC	AI - CSX	FI	AN	LEGA	UDEUR	Non Voto	
<i>Sa chi voterà</i>	<b>+12</b>	<b>0</b>	<b>-5</b>	<b>-2</b>	<b>+6</b>	<b>+7</b>	<b>+1</b>	<b>-2</b>	<b>-18</b>	121
<i>Area sì, partito no</i>	<b>-4</b>	<b>+1</b>	<b>+5</b>	<b>+2</b>	<b>-1</b>	<b>+3</b>	<b>-2</b>	<b>+1</b>	<b>-5</b>	67
<i>Indeciso</i>	<b>-1</b>	<b>0</b>	<b>+1</b>	<b>0</b>	<b>-3</b>	<b>-5</b>	<b>+1</b>	<b>+1</b>	<b>+6</b>	38
<i>Rifiuto</i>	<b>-7</b>	<b>-1</b>	<b>-1</b>	<b>0</b>	<b>-2</b>	<b>-5</b>	<b>0</b>	<b>0</b>	<b>+17</b>	45
	60	14	18	18	44	41	10	15	51	271

**Commento.** Le considerazioni fatte precedentemente sul PRC vengono confermate anche dagli altri scarti. In particolare, il PRC, rispetto alle altre formazioni politiche votate nel 2006, presenta lo scarto negativo più elevato (**-5**) in corrispondenza della modalità *Sa chi voterà*, compensato da un eguale scarto positivo (**+5**) sulla modalità *Area sì, partito no*. L'esito delle elezioni tenutesi a poche settimane dal sondaggio daranno ragione di questa previsione: il partito di Bertinotti sarà infatti fuori dal Parlamento. In generale, gli altri scarti indicano la tenuta dei partiti più grandi e uno scricchiolamento nel partito di Casini e Mastella che, come noto, eviterà per un pelo la stessa fine del PRC.

**Utilità e limiti nell'uso delle frequenze attese.** La neutralità dell'algoritmo, agente sul prodotto dei subtotali di riga e colonna normalizzato per N, consente la rappresentazione virtuale, **senza alcun vincolo di causalità**, dell'articolazione dei casi della ricerca secondo le modalità tipologiche prodotte dall'incrocio delle due classificazioni. Sta poi allo studioso attribuire un significato agli scarti (**fr - ft**) tra il valore reale e quello teorico di ogni cella.



**L'utilità** è dovuta all'immediatezza con cui si individuano le celle contenenti gli scarti positivi o negativi più consistenti.

**I limiti** sono così riassumibili:

- gli scarti ( **fr** – **ft** ) vanno calcolati manualmente;
- analoghi scarti, osservabili in celle della stessa riga o della stessa colonna, sono comparabili solamente attraverso la relativizzazione di ciascuno scarto, in *valore assoluto*, rispetto al sottotale di riga o colonna a cui vuole riferirsi.  
Per stare all'esercizio in corso, considerando i soggetti certi del loro voto alle imminenti elezioni politiche ( *Sa chi voterà* ), pesa di più lo scarto di -5 calcolato sul totale di 18 ex elettori di PRC (  $18/500=27,8\%$  ) piuttosto dello scarto di +7 calcolato sul totale di 44 ex elettori forzisti (  $44/700=15,9\%$  );
- gli scarti tra i valori reali e quelli attesi ( **fr** – **ft** ) mettono in luce l'effettiva ( spesso scarsa ) numerosità del campione. Nella tavola di contingenza oggetto dell'esercitazione, nonostante una popolazione di 271 casi, considerabile di modeste ma non piccole dimensioni, le 36 celle comportano l'inevitabile drastica riduzione delle frequenze in numerose di esse. Questo dato di fatto è una delle ragioni principali del mancato utilizzo delle frequenze teoriche attese da parte sia delle agenzie di sondaggi che della più vasta comunità di studiosi: la pochezza dei numeri rischia infatti di sminuire la valenza di *veritas* associata alle generalizzazioni teoriche evincibili dall'analisi dei dati. A tal proposito, basti pensare come sondaggi socio-politici di incontestabile valenza scientifica utilizzino campioni nazionali di 2000 o 3000 casi. Sono tanti, pur tuttavia anch'essi a rischio di disarticolazione a fronte di incroci tra classificazioni con numerose categorie.

Esistono altre due strade alternative nell'analisi delle **fr**. Esse prevedono, in relazione a specifiche necessità *descrittive* o *confermative*, il calcolo delle percentuali sul *totale dei casi* oppure *per riga* o *colonna*; queste ultime utilizzabili con lo stesso meccanismo logico interpretativo.

### C.9. ESERCIZIO n° 11. Crosstabs con percentuali sul totale [ *Total* ].

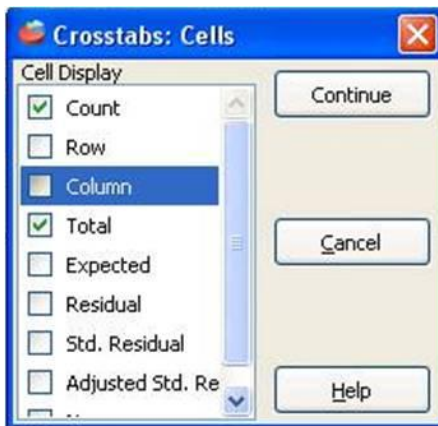
Il calcolo delle percentuali sul totale dei casi è un buon sistema per *descrivere* quanto ogni *fr* di cella *pesa* nel campione rispetto alle altre. Questa informazione è sicuramente utile ogniqualvolta le due classificazioni siano indipendenti tra loro e presentino differenti subtotali nelle loro categorie.

Un esempio intuitivo può essere quello relativo allo studio congiunto di due variabili di *background*, come la *Posizione rispetto al lavoro* e il *Genere*. L'obiettivo da perseguire diventa pertanto la descrizione della popolazione in relazione alla distribuzione dei casi secondo le modalità tipologiche prodotte dalle due variabili.

**ESERCIZIO.** Si richieda la tavola di contingenza, con il calcolo delle percentuali di cella sul totale, delle variabili D7 ( *Posizione rispetto al lavoro* ) e D3 ( *Genere* ).

**Esecuzione della procedura.** Eseguire i seguenti passi:

- Attivare la procedura **Analyze** ( → ) **Descriptive Statistics** ( → ) **Crosstabs** e, come al solito,
- evidenziare alternativamente *Posizione rispetto al lavoro* e *Genere*, spostando ( → ) la D7 e la D3 nei campi **Rows** e **Columns**. Dopodiché passare, in successione, nei campi delle funzioni **Statistics** e **Cells** ;



- nel campo **Statistics** deflaggare **Chsq** .
- Passare quindi alla funzione **Cells** ( qui illustrata a sinistra );
- deflaggare tutte le opzioni tranne la **Count** e
- flaggare l'opzione **Total** ;
- cliccare su *Continue* e tornare alla finestra iniziale del comando.
- Cliccare sul bottone **OK** per eseguire la procedura statistica.

In ciascuna delle 14 celle della tavola di contingenza compare, al di sotto del valore reale [ **count** ], la percentuale di *fr* sugli N casi [ **total** ].

Questo calcolo svincola ogni *fr* dal legame con i propri subtotali di riga e colonna, generando una quota che consente la comparazione del peso relativo di tutte le modalità tipologiche presenti nella tavola di contingenza, permettendo al tempo stesso di costruirne la gerarchia.

Nell'esempio specifico si potrà affermare che la prima

modalità tipologica più rappresentata è quella degli *uomini in attività lavorativa* ( 35,8% sul totale dei casi della ricerca ), seguita a distanza dall'analoga al femminile ( 17,34% ). Terzi e quarti, gli *studenti universitari* di sesso *maschile* ( 15,87% ) e *femminile* ( 9,23% ) . Al quinto posto, le *donne che studiano e lavorano* ( 5,9% ). Al sesto posto, gli *uomini che studiano e lavorano* ( 2,95% ). Le donne sopravanzano infine gli uomini nelle rimanenti categorie riferite alla *disoccupazione*, *ricerca della I occupazione*, essere in *pensione* e *casalinga*.

Questo è dunque il contributo descrittivo fornito dal calcolo della percentuale sul totale.

Posizione rispetto al lavoro	Genere		Total
	donna	uomo	
stud univers	25,00 9,23%	43,00 15,87%	68,00 25,09%
studio e lavoro	16,00 5,90%	8,00 2,95%	24,00 8,86%
disoccupato	5,00 1,85%	2,00 ,74%	7,00 2,58%
in cerca I occup	6,00 2,21%	3,00 1,11%	9,00 3,32%
lavoro	47,00 17,34%	97,00 35,79%	144,00 53,14%
pensione	7,00 2,58%	5,00 1,85%	12,00 4,43%
casalinga	7,00 2,58%	,00 ,00%	7,00 2,58%
Total	113,00 41,70%	150,00 58,30%	271,00 100,00%

### C.10. Percentuali di riga o colonna: strumenti efficaci nel controllo empirico di un modello asimmetrico di relazione tra due classificazioni.

Prima di sperimentare le altre varianti nell'esecuzione della procedura **Crosstabs**, si rende necessario fare il punto della situazione in ordine a quanto acquisito intorno all'uso delle tavole di contingenza.

Gli argomenti sviluppati finora indicano che:

- l'incrocio tra classificazioni è sempre possibile e genera una tavola a doppia entrata con un numero di celle (modalità tipologiche) pari al prodotto del numero di categorie delle due variabili;
- l'osservazione comparata del solo numero di casi conteggiabili all'interno di ciascuna cella (**fr**), in presenza di differenti subtotali di categoria, impedisce una corretta interpretazione della relazione tra le classificazioni;
- il calcolo delle percentuali delle **fr** di cella sul totale N può descrivere esclusivamente la gerarchia delle modalità tipologiche della tavola di contingenza, fornendo un contributo pressoché nullo all'individuazione di eventuali relazioni tra le due classificazioni;
- è necessario avere un quadro di riferimento teorico/statistico rispetto al quale confrontare la realtà effettiva rappresentata dalle **fr** conteggiate nelle celle;
- l'utilizzo delle frequenze teoriche attese (**ft**) e l'analisi dei loro scarti dai dati reali (**fr - ft**) rappresentano una soluzione valida del problema. Esse tuttavia, a causa dell'algoritmo utilizzato per calcolarle, oltre a svincolarsi dalla consistenza numerica dei subtotali di riga e colonna, rendono *simmetrico* il modello di relazione tra le classificazioni.

E' dunque necessario trovare soluzioni efficaci ai problemi sopra evidenziati, nella - supposta - eventualità in cui il modello di relazione preveda che una classificazione svolga la funzione di variabile *indipendente* e l'altra di variabile *dipendente*. A ben vedere, tale modello *asimmetrico* è presente nella stragrande maggioranza delle situazioni di ricerca, in cui lo studioso desidera confermare un'ipotesi: cioè il fatto di poter approssimativamente dimostrare che la numerosità presente in alcune modalità tipologiche (in eccesso o in difetto rispetto ad un quadro virtuale di reciproca indipendenza tra le classificazioni) è dovuta all'influenza di una variabile rispetto ad un'altra.

La tecnica in grado di risolvere la questione è quella del calcolo delle percentuali di *riga* o di *colonna* all'interno di ogni cella della tavola di contingenza.

**N.B. Le percentuali di *riga* o di *colonna*, dal punto di vista tecnico/statistico, vengono calcolate con il medesimo meccanismo, anche se i valori risultanti sono ovviamente differenti.**

Ciò che differenzia i due tipi di calcolo è il solo fatto di fare base a 100 il **subtotale di ogni riga** o il **subtotale di ogni colonna**.

	DS	MAR	PRC	AI - CSX	FI	AN	LEGA	UDEUR	Non Voto	
Sa chi voterà										121=100
Area sì, partito no										67=100
Indeciso										38=100
Rifiuto										45=100
	60=100	14=100	18=100	18=100	44=100	41=100	10=100	15=100	51=100	271

La scelta se calcolare le percentuali per *riga* o per *colonna* va fatta tenendo conto della natura *asimmetrica* del modello di relazione tra le classificazioni. In tal senso, si assume che **la variabile indipendente svolga un ruolo primario nel meccanismo di interpretazione del modello**. Poiché l'orientamento di lettura dei dati per *riga* o per *colonna* comporta - come già visto - specifici vincoli

interpretativi, diventa importante che lo studioso, nella definizione della tavola di contingenza, stabilisca il posizionamento della variabile *indipendente* [ in orizzontale ( *riga* ) o in verticale ( *colonna* ) ].

Riassumendo, il CRITERIO operativo da seguire nella scelta del tipo di percentuale ( per *riga* o *colonna* ) prevede:

- a. **Identificazione** della variabile *indipendente* ;
- b. Arbitraria **collocazione** della variabile *indipendente* in orizzontale ( *riga* ) o in verticale ( *colonna* ) [ Lo studioso sceglierà in base a preferenze estetiche o a regole tipografiche o altro ancora ].
- c. **Richiesta della percentuale di *riga* o *colonna* in base alla collocazione orizzontale o verticale scelta per la variabile indipendente.**

Date queste premesse, si sperimenterà la tecnica con il prossimo esercizio, con il quale si chiude la carrellata delle procedure bivariate previste da PSPP.

### C.10.1. ESERCIZIO n° 12: Crosstabs con % di colonna [ *Column* ].

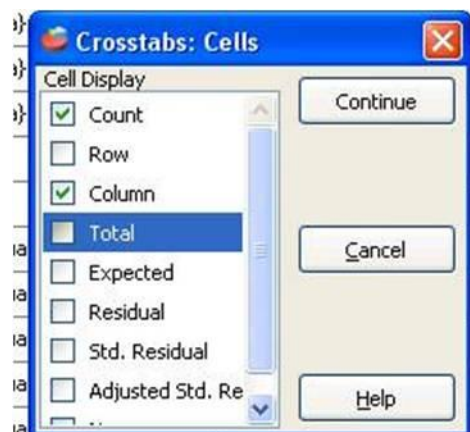
Eseguire la Crosstabs tra le variabili D13 e AggD2 con le percentuali di colonna [ *Column* ].

Da quanto argomentato nel capitolo precedente, la richiesta del calcolo delle percentuali di colonna è dovuto al fatto che:

- a. la variabile AggD2 svolge il ruolo di variabile *indipendente*, in qualità di indicatore dell'orientamento politico pregresso dei soggetti intervistati; mentre la D13 svolge il ruolo di variabile *dipendente*. Il modello di relazione assume pertanto che la sicurezza o l'indecisione nell'esprimere il proprio voto nelle imminenti elezioni 2008 sia influenzato dal giudizio sulle politiche svolte dai partiti votati nel 2006;
- b. si preferisce collocare la variabile AggD2 verticalmente ( *colonna* ), in modo tale da consentire una più semplice lettura del comportamento di voto fra gli ex elettori dei singoli partiti e una contestuale più semplice comparazione fra i partiti stessi.

**Esecuzione della procedura.** Si reiterano i passi già sperimentati nei due esercizi precedenti, e cioè:

- Attivare la procedura Analyze ( → ) Descriptive Statistics ( → ) Crosstabs ;
- evidenziare alternativamente *Comportamento alle elezioni 2008 e Partiti aggregati*, spostando ( → ) la D13 e la AggD2 nei campi Rows e Columns. Dopodiché ...



- nel campo Statistics deflaggare **Chsq** .
- Passare quindi alla funzione Cells ( qui illustrata a sinistra );
- deflaggare tutte le opzioni tranne la **Count** e
- flaggare l'opzione **Column** ;
- cliccare su *Continue* e tornare alla finestra iniziale del comando.
- Cliccare sul bottone OK per eseguire la procedura statistica.

**Esito :**

Comportamento alle elezioni 2008 \* Partiti aggregati [count, column %].

Comportamento alle elezioni 2008	Partiti aggregati									Total
	DS	MARGHERITA	PRC	ALTRO CSX	FI	AN	LEGA	UDEUR	Non voto	
sa chi voterà	39,00 65,00%	6,00 42,86%	3,00 16,67%	6,00 33,33%	26,00 59,09%	25,00 60,98%	6,00 60,00%	5,00 33,33%	5,00 9,80%	121,00 44,65%
area sì, partito no	11,00 18,33%	5,00 35,71%	9,00 50,00%	6,00 33,33%	10,00 22,73%	13,00 31,71%	,00 0,00%	5,00 33,33%	8,00 15,69%	67,00 24,72%
indeciso	7,00 11,67%	2,00 14,29%	4,00 22,22%	3,00 16,67%	3,00 6,82%	1,00 2,44%	2,00 20,00%	3,00 20,00%	13,00 25,49%	38,00 14,02%
rifiuto	3,00 5,00%	1,00 7,14%	2,00 11,11%	3,00 16,67%	5,00 11,36%	2,00 4,88%	2,00 20,00%	2,00 13,33%	25,00 49,02%	45,00 16,61%
Total	60,00 100,00%	14,00 100,00%	18,00 100,00%	18,00 100,00%	44,00 100,00%	41,00 100,00%	10,00 100,00%	15,00 100,00%	51,00 100,00%	271,00 100,00%

**Osservazioni.** La tavola di contingenza presenta le 36 fr e, al di sotto di ognuna, la percentuale calcolata ponendo a 100 il subtotalo di ogni colonna. In tal modo, la somma delle quattro percentuali nelle celle di ciascuna colonna sarà uguale al 100%.

**Criteri descrittivi di lettura della tavola di contingenza.** Sono sempre due gli orientamenti di lettura da utilizzare per l'interpretazione dei dati: per *riga* e per *colonna*. Come primo criterio di lettura verrà utilizzato quello conforme al tipo di percentuale di cella richiesta: quindi per *colonna*.

*Letture per colonna.* All'interno di ogni gruppo di casi distinto in base al partito votato nel 2006, si osserva la distribuzione percentuale delle quattro modalità di comportamento di voto dichiarato. Ciò consente di stimare la fedeltà interna del bacino elettorale di ciascun partito e il contestuale rischio di perdita del consenso, a causa dell'attrazione esercitata da un altro partito della stessa area politica, oppure della indecisione o del rifiuto di andare a votare.

*Letture per riga.* Poiché il calcolo percentuale per colonna elimina l'effetto dei differenti subtotali relativi ai partiti votati nel 2006, l'osservazione per *riga* consente la comparazione delle quote di una medesima modalità di comportamento di voto all'interno dei differenti bacini elettorali.

Ad es., osservando la categoria di risposta *sa chi voterà* (prima riga dell'incrocio) sarà possibile stilare la classifica dei partiti in base alla percentuale di riconferme: DS (65%), AN (61%), LEGA (60%), FI (59%), MARGHERITA (43%), UDEUR e ALTROCSX (33%), PRC (17%) e Non voto (10%). Tale operazione può ovviamente essere ripetuta anche sulle altre tre modalità di comportamento di voto.

**Individuazione dell'elemento base per la comparazione.** Le considerazioni appena fatte danno l'idea di come possono essere lette le fr e le loro percentuali. Occorre tuttavia trovare un criterio di confronto tra la rappresentazione della realtà e un'altra, virtuale, in grado di svelare eventuali relazioni non casuali tra le modalità congiunte delle due classificazioni. Nel far ciò, si dovrà tenere conto della natura *indipendente* della variabile AggD2 e *dipendente* della D13, superando lo schema *simmetrico* contemplato dalle frequenze teoriche attese (ft).

Il modello *asimmetrico* deve poter rispondere alla domanda "quali effetti più o meno accentuati sulle singole modalità di comportamento di voto (D13) sono imputabili ai partiti scelti nel 2006 (AggD2)?"

Il quesito impone quindi la lettura dei dati per *riga*, cioè secondo ogni modalità di comportamento di voto singolarmente presa.

L'**elemento base** in grado di consentire, per gli otto partiti e il gruppo degli astenuti, il calcolo degli scarti positivi o negativi sulla stessa modalità di comportamento di voto, sarà rappresentato dalla **percentuale di colonna calcolata sul totale di ogni singola modalità della variabile dipendente**.

Si considerino le percentuali di colonna (arrotondate) corrispondenti alla modalità *Sa chi voterà*.

	DS	MAR	PRC	AI - CSX	FI	AN	LEGA	UDEUR	Non Voto	% su 271
Sa chi voterà	65	43	17	33	59	61	60	33	10	45

Il valore **45** (percentuale arrotondata di 121 su 271) diventa l'elemento base da cui partire per evidenziare eventuali relazioni interpretabili all'interno delle nove modalità tipologiche. Esso rappresenta il valore statistico riferibile all'*ipotesi nulla*.

<b>IPOTESI NULLA</b>	DS	MAR	PRC	AI - CSX	FI	AN	LEGA	UDEUR	Non Voto	%
Sa chi voterà	<b>45</b>	<b>45</b>	<b>45</b>	<b>45</b>	<b>45</b>	<b>45</b>	<b>45</b>	<b>45</b>	<b>45</b>	<b>45</b>

Qualora l'*ipotesi nulla* fosse confermata, il legame tra gli effetti sul *sapere chi votare* e il *partito scelto* due anni prima sarebbe dovuto totalmente al caso. Le percentuali di colonna calcolate in ogni singola cella della riga risulterebbero quindi uguali a quella calcolata sul totale, e le nove differenze **fr-45** darebbero risultato 0.

Nello specifico dell'esercizio fatto, l'*ipotesi nulla* viene smentita. I nove scarti percentuali **fr-45** sono infatti tutti diversi da zero: quattro positivi e cinque negativi.

	DS	MAR	PRC	AI - CSX	FI	AN	LEGA	UDEUR	Non Voto	%
Sa chi voterà	65-45= <b>20</b>	43-45= <b>-2</b>	17-45= <b>-38</b>	33-45= <b>-12</b>	59-45= <b>14</b>	61-45= <b>16</b>	60-45= <b>15</b>	33-45= <b>-12</b>	10-45= <b>-35</b>	<b>45</b>

Il meccanismo di interpretazione parte dall'elemento base, costituito dalla percentuale di soggetti che hanno dichiarato di *sapere per chi votare*. Si dirà pertanto:

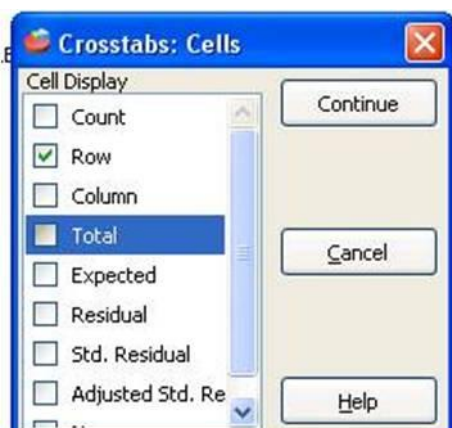
rispetto al 45% di intervistati che hanno dichiarato di *sapere chi voteranno alle prossime elezioni*, la quota di chi aveva votato PRC è di solo il 17%. La differenza tra questa percentuale e quella sul totale indica una perdita proporzionale di 38 punti, prefigurante un disastroso calo di consensi. La previsione si

consolida osservando gli scarti relativi alle altre modalità di comportamento di voto: sono infatti proporzionalmente maggiori gli *indecisi* e, soprattutto, coloro che *sceglieranno un altro partito all'interno della stessa area*.

	PRC	%
Sa chi voterà	17-45= <b>-38%</b>	<b>45</b>
Area sì, partito no	50-25= <b>25%</b>	<b>25</b>
Indeciso	22-14= <b>8%</b>	<b>14</b>
Rifiuto	11-16= <b>-5%</b>	<b>16</b>

Questi sono dunque i suggerimenti per impostare correttamente una tavola di contingenza ogniquale volta l'incrocio delle due classificazioni sia contemplabile all'interno di un modello di relazione causa → effetto.

La scelta delle percentuali di riga o colonna è, come già detto, legata al posizionamento verticale (colonna) od orizzontale (riga) della variabile indipendente. Il criterio di utilizzo delle percentuali per *riga* è, alla fine, identico a quello visto per le percentuali di colonna. Nel caso specifico dell'esempio fatto, variando il posizionamento della AggD2 e chiedendo, conseguentemente, il solo calcolo delle percentuali per riga, otterremo esattamente gli stessi valori percentuali già visti sopra.



Partiti aggregati	Comportamento alle elezioni 2008				Total
	sa chi voterà	area sì, partito no	indeciso	rifiuto	
DS	65,00%	18,33%	11,67%	5,00%	100,00%
MARGHERITA	42,86%	35,71%	14,29%	7,14%	100,00%
PRC	16,67%	50,00%	22,22%	11,11%	100,00%
ALTRO CSX	33,33%	33,33%	16,67%	16,67%	100,00%
FI	59,09%	22,73%	6,82%	11,36%	100,00%
AN	60,98%	31,71%	2,44%	4,88%	100,00%
LEGA	60,00%	,00%	20,00%	20,00%	100,00%
UDEUR	33,33%	33,33%	20,00%	13,33%	100,00%
Non voto	9,80%	15,69%	25,49%	49,02%	100,00%
Total	44,65%	24,72%	14,02%	16,61%	100,00%