

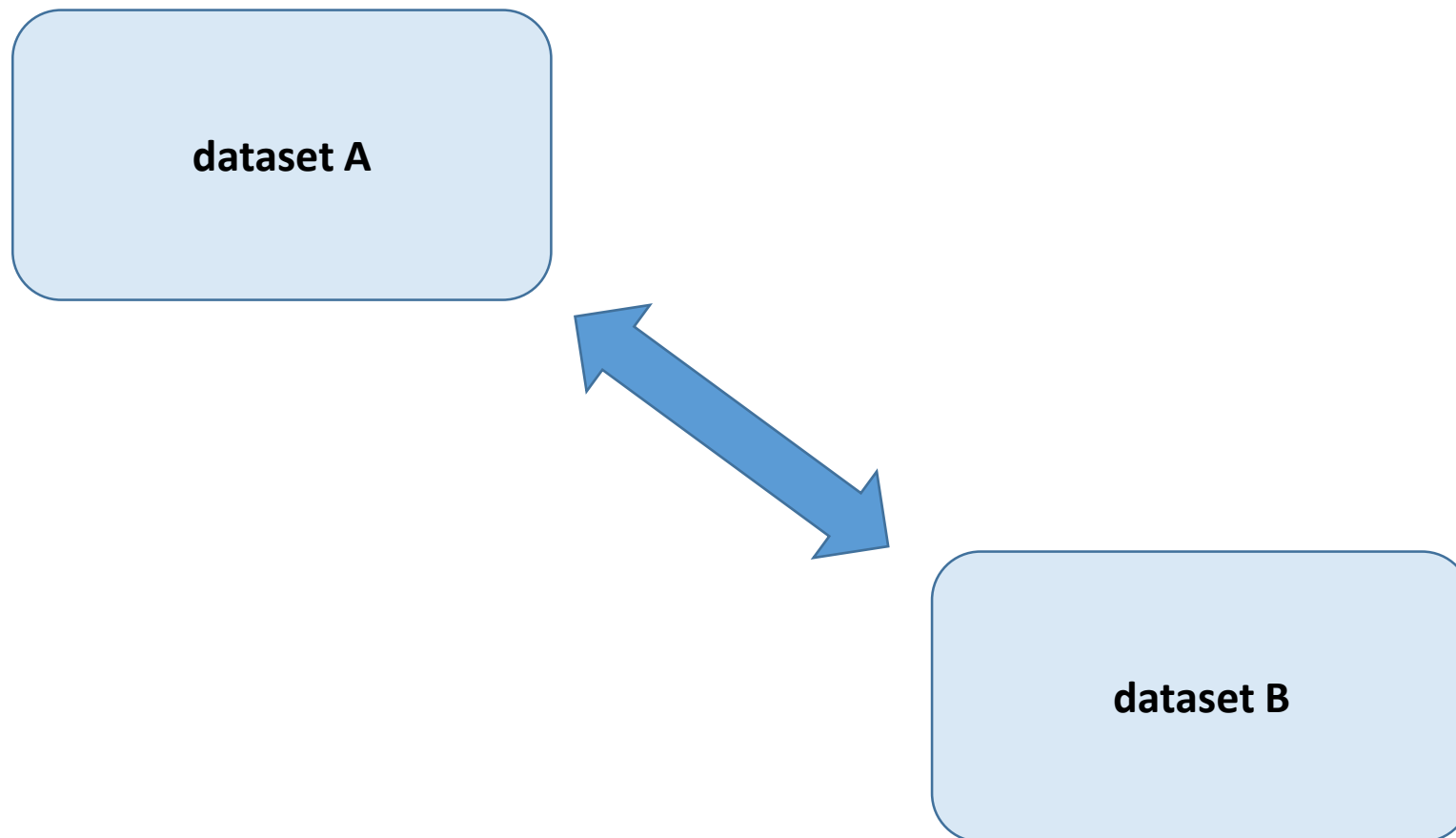
Programmazione avanzata in SAS

Lo statement MERGE

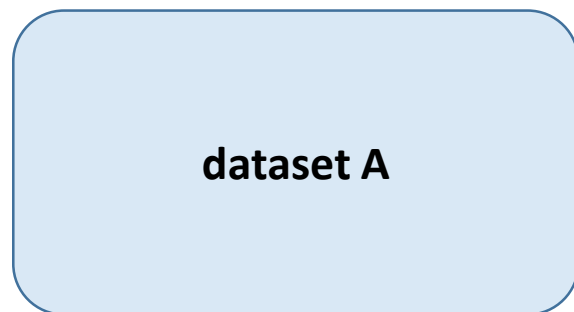
Denis QUARTA

5 febbraio 2020

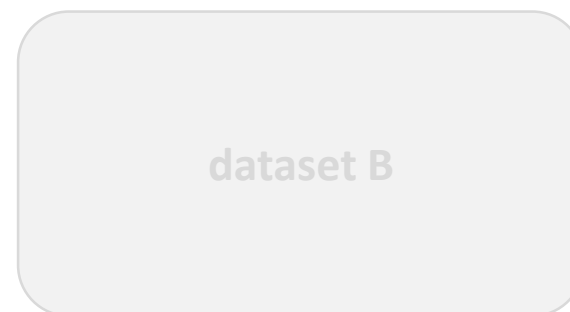
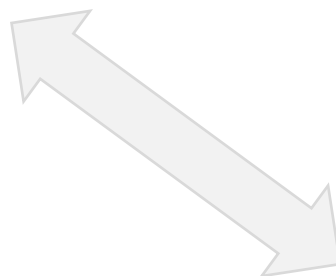
MERGE



MERGE



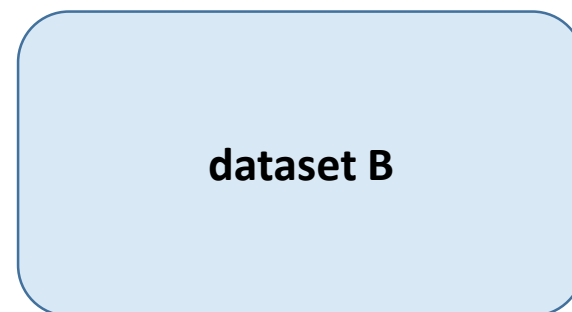
X osservazioni
P variabili



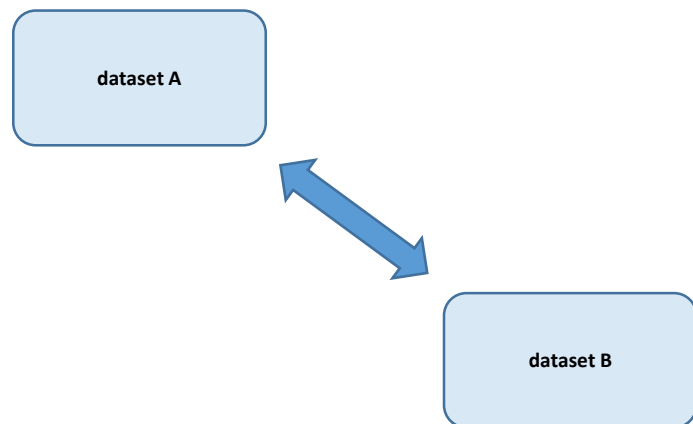
MERGE



Y osservazioni
Q variabili



MERGE



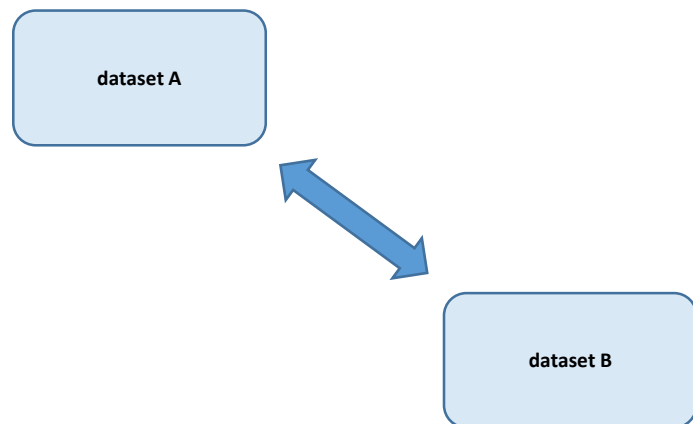
Condizione necessaria (1):

Fra
le P variabili di A e
le Q variabili di B

Ve ne sia almeno una che *misuri* la stessa caratteristica (es. CODFIS, MATRICOLA, ID) e che abbia ugual nome e la stessa struttura in entrambi i dataset.

Questa variabile, viene definita quale chiave di linkage

MERGE



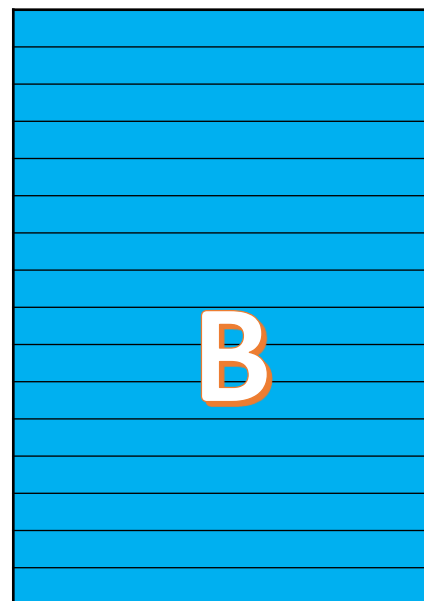
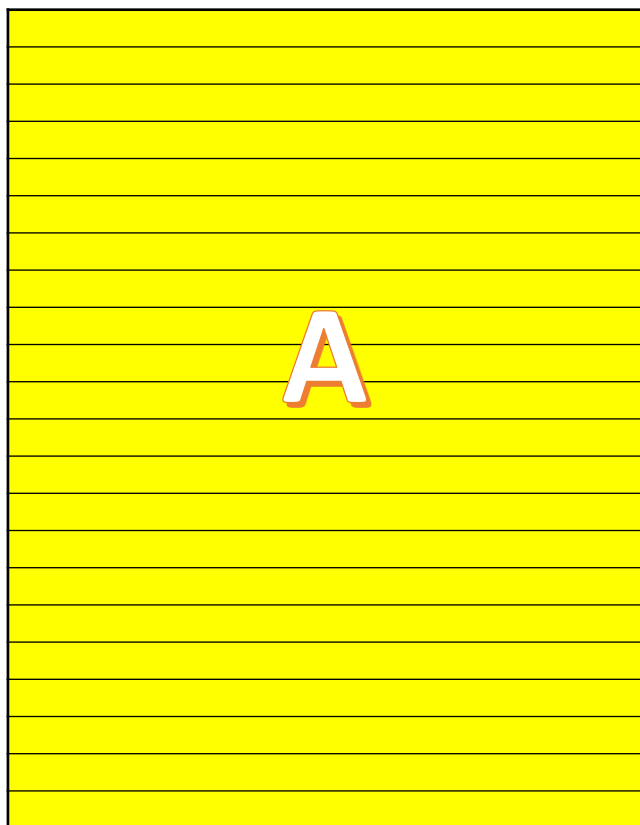
Condizione necessaria (2):

La variabile utilizzata come chiave di linkage deve esser precedentemente ordinata sia nel dataset A, sia nel dataset B.

Si utilizzerà una procedura di ordinamento – proc sort.

MERGE

Tecniche di LINKAGE – Inner (1)



X osservazioni in A
Y osservazioni in B

[illegible]

MERGE

Tecniche di LINKAGE – Inner (3)

A1	A2	A3	A4	A5	ID	ID	B2	B3	B4
					123	109			
					126	551			
					555	126			
					556	101			
					456	154			
					654	159			
					899	973			
					788	379			
					211	612			
					456	621			
					951	211			
					417	258			
					154	321			
					321	761			
					417	700			
					258				
					852				
					963				
					159				
					951				
					111				

ID non è ordinata né in A né in B

MERGE

Tecniche di LINKAGE – Inner (4)

A1	A2	A3	A4	A5	ID	ID	B2	B3	B4
					111	101			
					123	109			
					126	126			
					154	154			
					159	159			
					211	211			
					258	258			
					321	321			
					417	379			
					417	551			
					456	612			
					456	621			
					555	700			
					556	761			
					654	973			
					788				
					852				
					899				
					951				
					951				
					963				

INNER JOIN

Unisce solo le righe dove la chiave corrisponde.

Elimina le altre.

MERGE

Tecniche di LINKAGE – Inner (5)

A1	A2	A3	A4	A5	ID	ID	B2	B3	B4
					111	101			
					123	109			
					126	126			
					154	154			
					159	159			
					211	211			
					258	258			
					321	321			
					417	379			
					417	551			
					456	612			
					456	621			
					555	700			
					556	761			
					654	973			
					788				
					852				
					899				
					951				
					951				
					963				

Crea un nuovo dataset C, prendendo tutte le variabili di A e di B.
La chiave, viene presa una sola volta

MERGE

Tecniche di LINKAGE – Inner (6)

A1	A2	A3	A4	A5	ID	B2	B3	B4
					111			
					123			
					126			
					154			
					159			
					211			
					258			
					321			
					417			
					417			
					456			
					456			
					555			
					556			
					654			
					788			
					852			
					899			
					951			
					951			
					963			

Quante osservazioni?

X?

Y?

Tecniche di LINKAGE – Inner (7)

A1	A2	A3	A4	A5	ID	B2	B3	B4
					126			
					154			
		C			159			
					211			
					258			
					321			

Tecniche di LINKAGE – Inner (2)

```
data B;  
set B;  
rename b1=id;  
run;
```

[illegible]

5 febbraio 2020

MERGE

Tecniche di LINKAGE – SAS (2)

```
proc sort data=A;  
by id;  
run;
```

```
proc sort data=B;  
by id;  
run;
```

Tecniche di LINKAGE – Inner (3)

A1	A2	A3	A4	A5	ID	ID	B2	B3	B4
					123	109			
					126	551			
					555	126			
					556	101			
					456	154			
					654	159			
					899	973			
					788	379			
					211	612			
					456	621			
					951	211			
					417	258			
					154	321			
					321	761			
					417	700			
					258				
					852				
					963				
					159				
					951				
					111				

ID non è ordinata né in A né in B

MERGE

Tecniche di LINKAGE – SAS (3)

```
data c;  
merge a (in=p) b (in=q) ;  
by id;  
if p and q;  
run;
```

dichiaro il nome del nuovo dataset

dichiaro quali dataset voglio linkare

definisco la chiave di linkage (presente in entrambi i dataset)

scelgo un inner join

Tecniche di LINKAGE – Inner (5)

A1	A2	A3	A4	A5	ID	ID	B2	B3	B4
					111	101			
					123	109			
					126	126			
					154	154			
					159	159			
					211	211			
					258	258			
					321	321			
					417	379			
					417	351			
					456	612			
					456	621			
					555	700			
					556	761			
					654	973			
					788				
					852				
					899				
					951				
					951				
					963				

Crea un nuovo dataset C, prendendo tutte le variabili di A e di B. La chiave, viene presa una sola volta

Tecniche di LINKAGE – Numero delle osservazioni

Chiave di linkage è univoca o si ripete?

UNIVOCA – Linkage 1 a 1

SI RIPETE, solo in un dataset

Linkage $n*1$

SI RIPETE, in tutti i dataset

Linkage $n*m$

NON SUPPORTATO DAL MERGE

Tecniche di LINKAGE – Numero delle osservazioni

Chiave di linkage è univoca o si ripete?

UNIVOCA – Linkage 1 a 1

SI RIPETE, solo in un dataset
Linkage n*1

SI RIPETE, in tutti i dataset
Linkage n*m

NON SUPPORTATO DAL MERGE

```
217  
218 data c;  
219 merge a (in=p) b (in=q);  
220 by id;  
221 if p and q ;  
222 run;
```

NOTE: There were 20 observations read from the data set WORK.A.

NOTE: There were 14 observations read from the data set WORK.B.

NOTE: The data set WORK.C has 6 observations and 9 variables.

NOTE: DATA statement used (Total process time):

real time	0.04 seconds
cpu time	0.00 seconds

MERGE

Tecniche di LINKAGE – Numero delle osservazioni

Chiave di linkage è univoca o si ripete?

UNIVOCA – Linkage 1 a 1

```
231 data c;  
232 merge a (in=p) b (in=q);  
233 by id;  
234 if p and q ;  
235 run;
```

SI RIPETE, solo in un dataset
Linkage n*1

```
NOTE: There were 20 observations read from the data set WORK.A.  
NOTE: There were 14 observations read from the data set WORK.B.  
NOTE: The data set WORK.C has 8 observations and 9 variables.  
NOTE: DATA statement used (Total process time):  
      real time           0.07 seconds  
      cpu time            0.01 seconds
```

SI RIPETE, in tutti i dataset
Linkage n*m

NON SUPPORTATO DAL MERGE

	A1	A2	A3	A4	A5	id	B2	B3	B4
1	pippo	minnie	giallo	100	n	126	more	dom	0
2	pippo	topolino	verde	10	n	154	lamponi	dom	0
3	pippo	minnie	arancio	10	n	159	more	dom	0
4	pippo	topolino	blu	10	n	211	fragole	dom	0
5	pippo	minnie	viola	10	n	258	lamponi	dom	0
6	pippo	topolino	grigio	100	n	321	lamponi	dom	1
7	pippo	minnie	azzurro	10	n	417	more	dom	1
8	pippo	topolino	verde	10	n	417	more	dom	1

Tecniche di LINKAGE – Numero delle osservazioni

Chiave di linkage è univoca o si ripete?

UNIVOCA – Linkage 1 a 1

SI RIPETE, solo in un dataset
Linkage $n*1$

SI RIPETE, in tutti i dataset
Linkage $n*m$

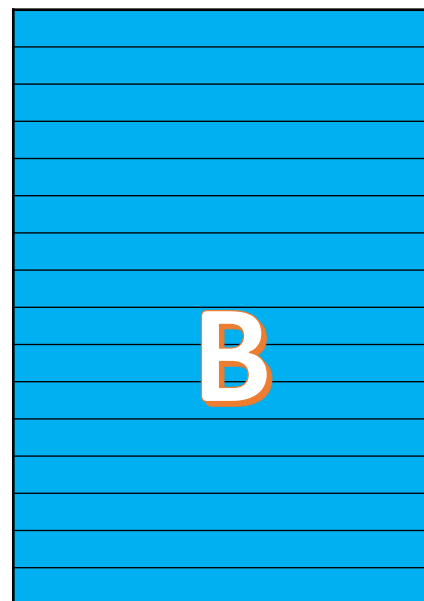
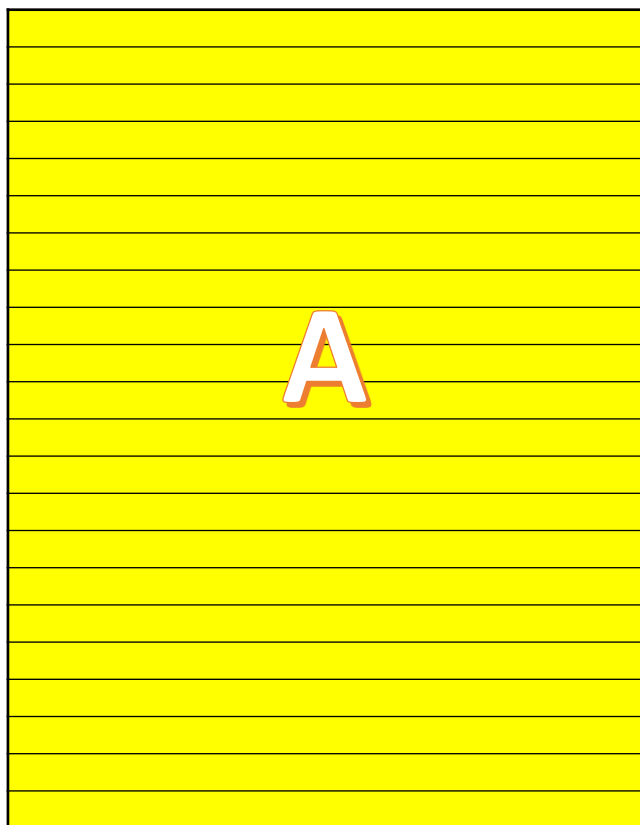
NON SUPPORTATO DAL MERGE

Produce un $n+m$.

Necessaria un'altra procedura all'interno del
linguaggio SQL.

MERGE

Tecniche di LINKAGE – Left (1)



X osservazioni in A
Y osservazioni in B

**EPIDEMIOLOGIA
PIEMONTE**

[illegible]

MERGE

Tecniche di LINKAGE – Left (3)

A1	A2	A3	A4	A5	ID	ID	B2	B3	B4
					123	109			
					126	551			
					555	127			
					556	101			
					456	154			
					654	159			
					899	973			
					788	379			
					211	612			
					456	621			
					951	211			
					417	258			
					154	321			
					321	761			
					417	700			
					258				
					852				
					963				
					159				
					951				
					111				

ID non è ordinata né in A né in B

MERGE

Tecniche di LINKAGE – Left (4)

A1	A2	A3	A4	A5	ID	ID	B2	B3	B4
					111	101			
					123	109			
					126	127			
					154	154			
					159	159			
					211	211			
					258	258			
					321	321			
					417	379			
					417	551			
					456	612			
					456	621			
					555	700			
					556	761			
					654	973			
					788				
					852				
					899				
					951				
					951				
					963				

LEFT JOIN

Prende tutte le righe di A, e dove la chiave corrisponde aggiunge le variabili di B.

Nessuna riga di A è eliminata.

MERGE

Tecniche di LINKAGE – Left (5)

A1	A2	A3	A4	A5	ID	ID	B2	B3	B4
					111	101			
					126	127			
					154	154			
					159	159			
					211	211			
					258	258			
					321	321			
					417	379			
					417	551			
					456	612			
					456	621			
					555	700			
					556	761			
					654	973			
					788				
					852				
					899				
					951				
					951				
					963				

Crea un nuovo dataset C, prendendo tutte le variabili di A e di B.

La chiave, viene presa una sola volta dall'archivio A

MERGE

Tecniche di LINKAGE – Left (6)

A1	A2	A3	A4	A5	ID	B2	B3	B4
					111			
					126			
					154			
					159			
					211			
					258			
					321			
					417			
					417			
					456			
					456			
					555			
					556			
					654			
					788			
					852			
					899			
					951			
					951			
					963			

C

Quando la chiave di A non trova corrispondenza in B, le variabili risultano non compilate

**EPIDEMIOLOGIA
PIEMONTE**

Tecniche di LINKAGE – Left (6)

scelgo un left join

Quando la chiave di A non trova corrispondenza in B, le variabili risultano non compilate



Denis QUARTA

m. denis.quarta@epi.piemonte.it
denis.quarta@unito.it

t. 01140188206

www.epi.piemonte.it