

Dati multi-fonte e analisi territoriali

Prof. Marco Tosi

Dipartimento di Scienze Statistiche

Fonti multiple e dati territoriali: un esempio

- Obiettivo: descrivere la distanza geografica/abitativa tra parenti nelle famiglie italiane (genitori, figli, e fratelli).

- **Dati Famiglie e Soggetti Sociali 2016**  **Istat** | Istituto Nazionale di Statistica
 - Campione di circa 25000 famiglie in cui viene intervistato un membro di 18+ selezionato in modo casuale.
 - Campione selezionato: 6758 intervistati tra i 50 e 65 con almeno un figlio, genitore o fratello in vita.
 - Escludiamo coloro che hanno solo parenti (nella maggioranza dei casi sono figli) che vivono nella stessa famiglia.
 - > **Scelta del campione ha implicazioni su potenziali distorsioni**
Heckman (1977) sample selection bias as a specification error

Dati relazionali

- Possiamo creare dati relazionali: Ego/ rispondente -> 3 figli max.
- -> 3 fratelli max.
- -> 2 genitori max.
- Dati relazionali analizzati come gerarchici dove:
 - Livello 1: relazione Ego-parente (20911).
 - Livello 2: Ego (6758).
 - Livello 3: Provincia dove abita Ego durante l'intervista (105 province con almeno 10 rispondenti).

3.4 ^{105 311} Può dirci se sono maschi o femmine, che età hanno e dove abitano?
(Indicarlo per ciascun fratello o sorella che non vive con Lei. Se ha più di 3 fratelli o sorelle che non vivono con Lei, riferirsi ai 3 che abitano più vicino)

COMPILARE PER COLONNA

	1° fratello/ sorella	2° fratello/ sorella	3° fratello/ sorella
Sesso:			
Maschio	1 ☐	1 ☐	1 ☐
Femmina	2 ☐	2 ☐	2 ☐
Età	1 ☐ 2 ☐ 3 ☐	1 ☐ 2 ☐ 3 ☐	1 ☐ 2 ☐ 3 ☐
Dove abitano:			
In un altro appartamento dello stesso caseggiato	1 ☐	1 ☐	1 ☐
Nello stesso Comune:			
Entro 1 km	2 ☐	2 ☐	2 ☐
Nel resto del Comune	3 ☐	3 ☐	3 ☐
In un altro Comune in Italia distante:			
Meno di 16 km	4 ☐	4 ☐	4 ☐
Da 16 a 50 km	5 ☐	5 ☐	5 ☐
Più di 50 Km	6 ☐	6 ☐	6 ☐
All'estero	7 ☐	7 ☐	7 ☐

Modello gerarchico

- Vivere a meno di 15km da genitori, figli e fratelli.

$$\begin{aligned} & \text{Logit}\{\text{Pr.}(\text{Prox}_{drp} \leq 15)\} \\ &= \beta_0 + \beta_1(\text{KinT}, \text{KAge}, \text{KGender})_{drp} + \beta_2(\text{KinN}, \text{Age}, \text{Gender})_{rp} \\ &+ \beta_3(\text{Density}, \text{Unemp}, \text{Norms})_p + \mu_{rp} + \nu_p + \nu_{1p}(\text{KinT})_{drp} \end{aligned}$$

- Modelliamo il tipo di relazione (*KinT*: genitore, figlio o fratello) in quanto i legami orizzontali sono meno vincolati normativamente.
- Random slope: ***KinT*** per modellare come essere genitore, figlio o fratello varia a seconda della provincia.
- ***Density e Unemp*** (tassi di disoccupazione) da **fonte Censimento 2011**, **Norms** include aspetti valoriali presi dallo stesso questionario (media provinciale).

Modello gerarchico: Risultati

Tuttavia la variabilità
esplicitata dal Random
Slope è poco interpretabile
in termini sostanziali:
**vivono più vicino al sud o
al nord, est o ovest?**

	Model 4	
	Coef.	S.E.
Type of kin (ref. Children)		
Sibling	-0.43**	(0.12)
Parents	0.12	(0.22)
Gender of anchor	-0.29**	(0.07)
Gender of kin	0.04	(0.04)
Age of anchor	0.30	(0.21)
Age of anchor^2	-0.00	(0.00)
Age of kin	0.01*	(0.00)
N. of children (Ref. 0)		
1	0.01	(0.12)
2	0.02	(0.11)
3+	-0.14	(0.13)
N. of siblings (Ref. 0)		
1	-0.08	(0.16)
2	-0.41*	(0.16)
3+	-0.48*	(0.16)
N. of parent (Ref. 0)		
1	0.15+	(0.08)
2	0.00	(0.11)
Province-level indicators:		
Population density per 100 km ²	-0.00	(0.01)
Traditional family norms	0.47	(0.36)
Unemployment rates	0.05**	(0.01)
Constant	-9.23	(6.06)
Random Parameters at individual level:		
Intercept variance	4.73**	(0.22)
Random Parameters at province level:		
Intercept variance	0.17**	(0.06)
Slope variance for Sibling	0.22**	(0.05)
Slope variance for Parents	0.10	(0.05)
N. of anchor-kin dyads	20,911	
N. of anchor-respondents	6,403	
N. provinces	105	

KinT a livello 1 di relazione

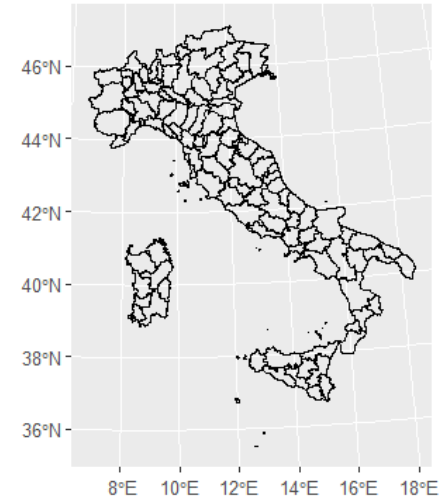
Variabili macro a livello di provincia
Che dovrebbero spiegare la variabilità di
KinT a livello 3

Varianza KinT a livello 3 di provincia

Come rappresentare questi risultati?

```
# Shapefile of nuts-3 regions from ISTAT o EUROSTAT
```

```
> nuts_i <- st_read("G:/Il mio Drive/PRIN PNRR/KIN Geography IT/Data & codes/Limiti01012010_g/Prov01012010_g")  
ggplot() + + geom_sf(data = nuts_i, color = "black", size = 0.6, fill= NA)
```



```
# Attributes: attributi che vogliamo plottare, ossia prob. predette
```

```
map_attr <- read_excel("G:/Il mio Drive/PRIN PNRR/KIN Geography IT/Data & codes/province2.xls")
```

```
# Merge: record linkage deterministico
```

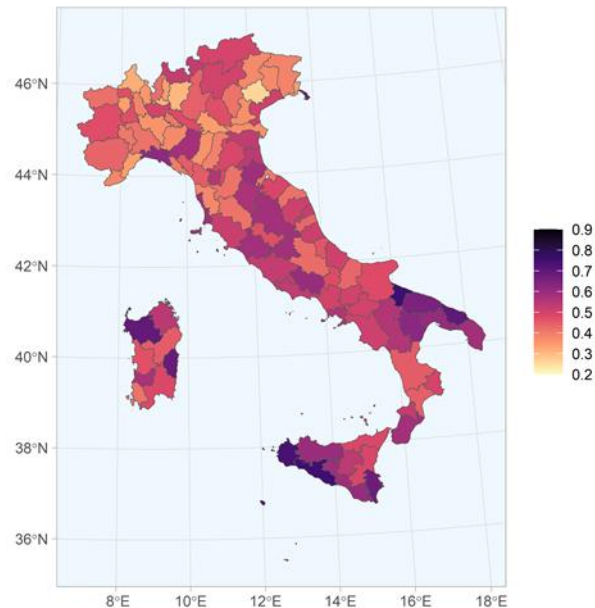
```
nuts3.df <- merge(nuts_i, map_attr, by = "SIGLA")
```

```
# Plot
```

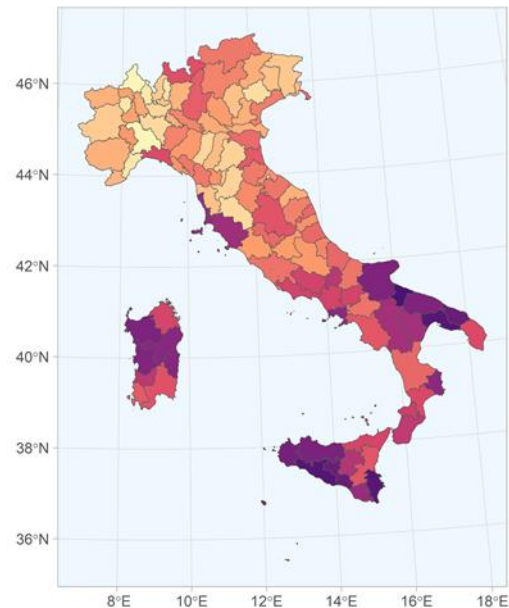
```
ggplot() + + geom_sf(data = nuts3.df, mapping = aes(fill = mean_par1), size = .2) +  
  + scale_fill_viridis_c(option = "A", limits=c(0.2, 0.9), breaks=seq(0.2,0.9,by=0.1),  
  + direction = -1, guide= "colourbar", na.value = "gray") +  
  + theme_light()+  
+ theme(legend.title = element_blank(), + panel.background = element_rect(fill = "aliceblue"))
```

Probabilità predette di vivere vicino

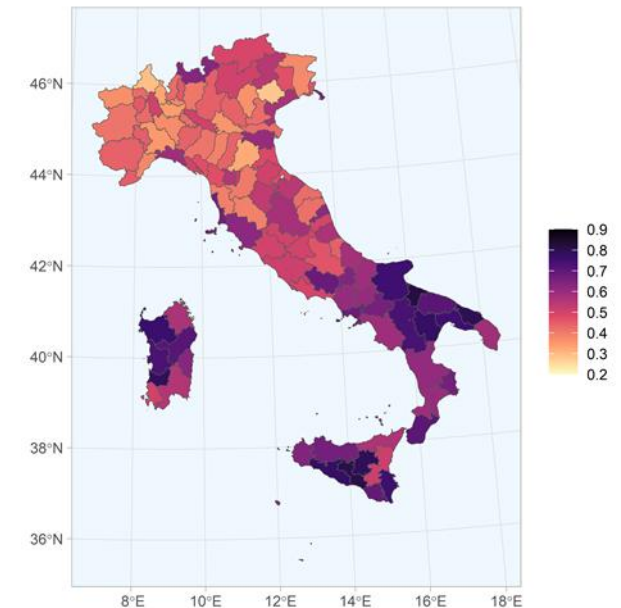
Figli



Fratelli



Genitori



Probabilità predette stimate dal modello in cui non sono incluse la variabili macro, ossia densità, disoccupazione e norme familiari a livello provinciale

Presentazione della BASE DATI Easy SHARE

Codici e Dati su Moodle

Presentazione della BASE DATI SHARE

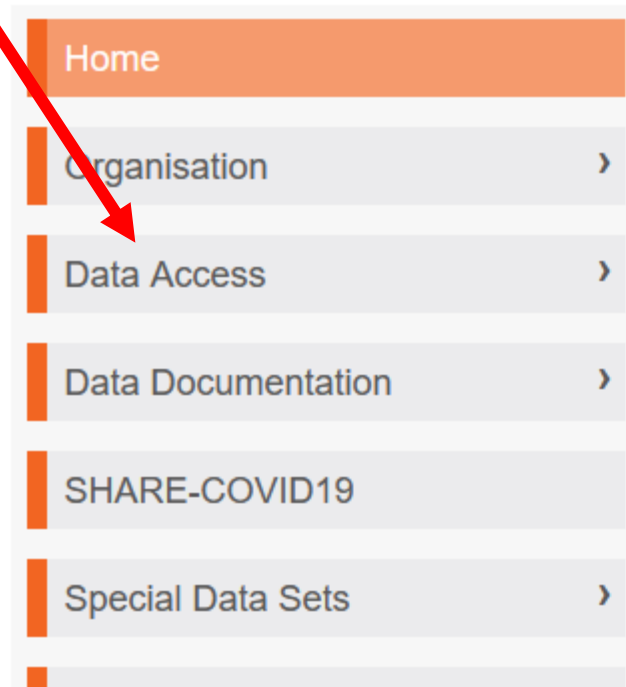
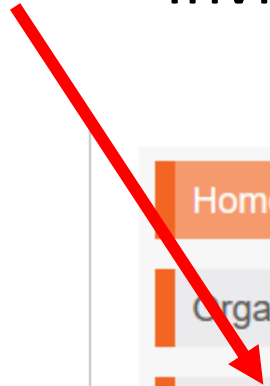
- WEB: <http://www.share-project.org/home0.html>



Survey of Health,
Ageing and
Retirement in
Europe

Iscrizione per l'utilizzo dei dati

- Compilare il modulo che trovate anche su moodle
- Inviare il modulo a: share-rdc@centerdata.nl



You are here: Home

SHARE - Survey of Health, Ageing and Retirement in Europe

The Survey of Health, Ageing and Retirement in Europe (SHARE) is a research infrastructure for studying the effects of health, social, economic and environmental policies over the life-course of European citizens and beyond. From 2004 until today, 480,000 in-depth interviews with 140,000 people aged 50 or older from 28 European countries and Israel have been conducted. Thus, SHARE is the largest pan-European social science panel study providing internationally comparable longitudinal micro data which allow insights in the fields of public health and socio-economic living conditions of European individuals.

Registrazione

- Gli utenti si devono registrare per poi scaricare gratuitamente i dati.
- Compilare lo statement per registrarsi.



STATEMENT CONCERNING THE USE OF SHARE DATA

To apply for access to the SHARE data, applicants must complete the following form.

Personal and professional details *(if handwritten: please use capital letters)*

Last name, first name/s:

Email address:

Position and job title:

Institution/organisation name:

- Type of institution/organisation:
(please tick appropriate option)
- ☐ University
 - ☐ Scientific research institution
 - ☐ Clearly separate and independent research dept. of a public institution/non-profit organisation
 - ☐ Other, **non**-scientific affiliation

→ *Note: Only if this option applies in your case you have to complete **PAGE 2** of this statement*

Data download

Current Scientific Releases of SHARE Data

Release	Stata	SPSS	R
Wave 8 COVID-19 Survey Release 1.0.0 DOI: 10.6103/SHARE.w8ca.100 2021-06-23 12:54:03	 sharew8_rel1-0-0_ca_ALL_datasets_stata.zip (8.83MB)	 sharew8_rel1-0-0_ca_ALL_datasets_spss.zip (8.85MB)	No R file available
Wave 8 Release 1.0.0 DOI: 10.6103/SHARE.w8.100 2021-06-23 12:52:24	 sharew8_rel1-0-0_ALL_datasets_stata.zip (44.83MB)	 sharew8_rel1-0-0_ALL_datasets_spss.zip (47.95MB)	No R file available

Dati SHARE- Survey of Health, Ageing and Retirement in Europe

- Quali Scopi:

“Health, social, economic and environmental policies over the life-course of European citizens and beyond. From **2004 until today, 480,000 in-depth interviews with 140,000 people aged 50 or older from 28 European countries and Israel** have been conducted. Thus, SHARE is the largest pan-European social science **panel study** providing internationally comparable longitudinal micro data which allow insights in the fields of public health and socio-economic living conditions of European individuals”.

Eligibility rules

- The SHARE target population consists of **all persons aged 50 years and over** at the time of sampling who have their regular domicile in the respective SHARE country. Persons are excluded if they are **incarcerated, hospitalised or out of the country during the entire survey period, unable to speak the country's language(s) or have moved to an unknown address**. In addition – in all waves – current **partners living in the same household are interviewed regardless of their age**.
- All SHARE respondents who were interviewed in any previous wave are part of the **longitudinal sample**. If they have a **new partner** living in the household, the new partner is eligible for an interview as well, regardless of age. Age eligible respondents who participated are traced and re-interviewed if they move within the country and end-of-life interviews are conducted if they decease. Younger partners, new partners and partners who never participated in SHARE will not be traced if they move and are not eligible for an end-of-life interview.

Data collection

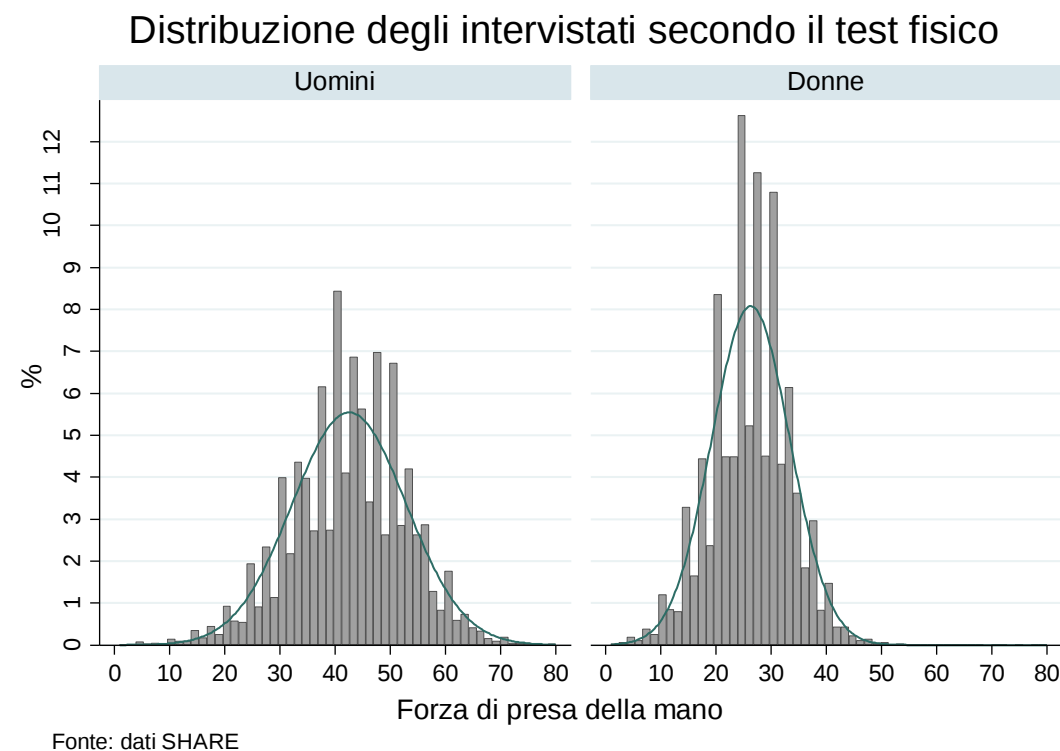
- SHARE data collection is based on **computer-assisted personal interviewing (CAPI)**. Personal interviews are necessary for SHARE because they make the execution of **physical tests possible**.
- There is one common **generic questionnaire** that is translated into the national languages (in some countries more than one language) using an internet based translation tool and processed automatically in a CAPI instrument. However, some internationally highly diverse variables require **country-specific measurements** and ex-post harmonisation, for example in the areas of education (ISCED) or occupation (ISCO, NACE).

Contenuti – Test fisici – un esempio

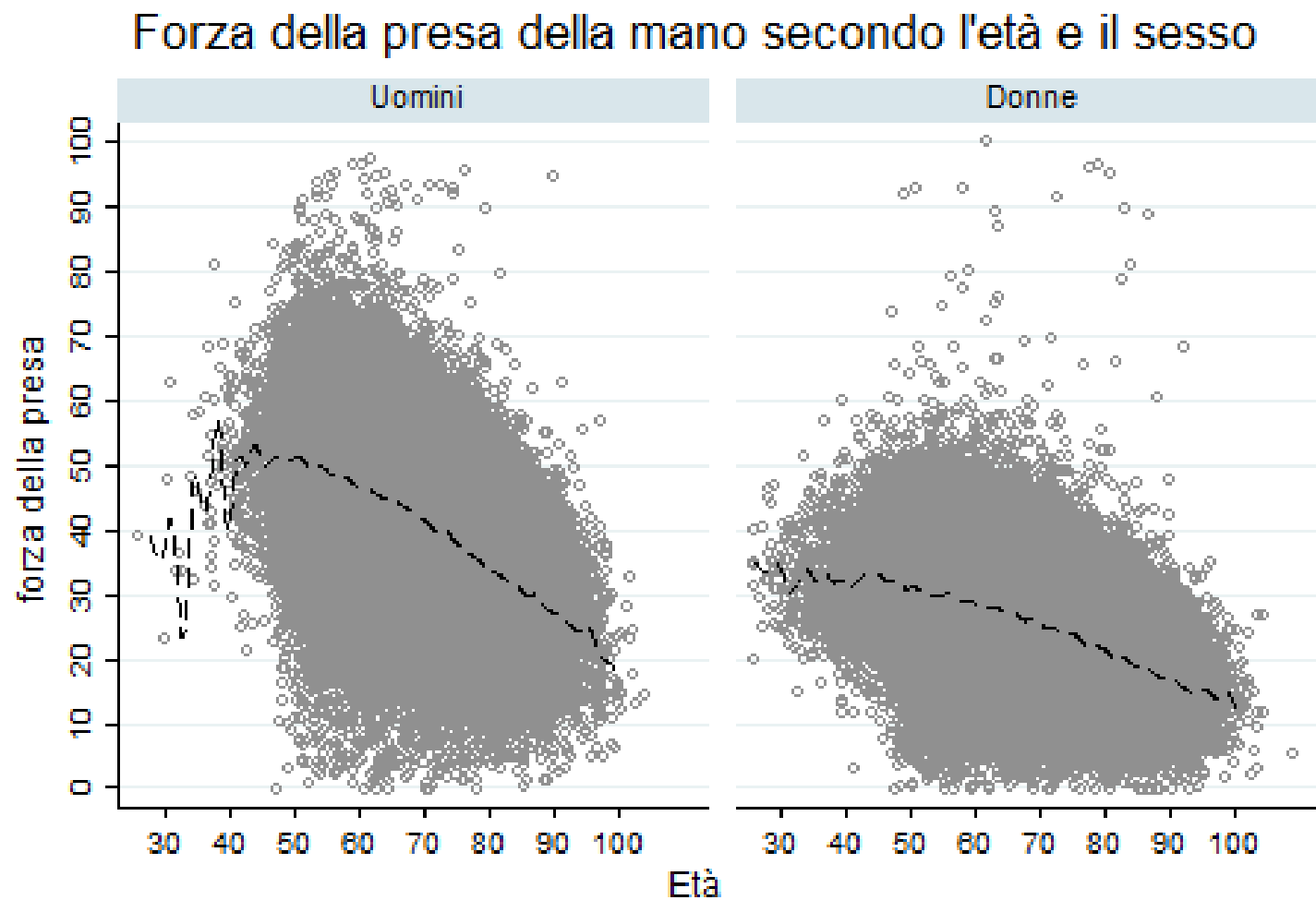
- Test fisici durante l'intervista:
- ***Dinamometro*** per la forza di presa della mano (Grip Strength).

La stretta di mano può costituire un metodo semplice ed economico per valutare il rischio individuale di morte e malattie cardiovascolari.

Unità di misura: *Kg di forza della presa*



Forza della presa della mano



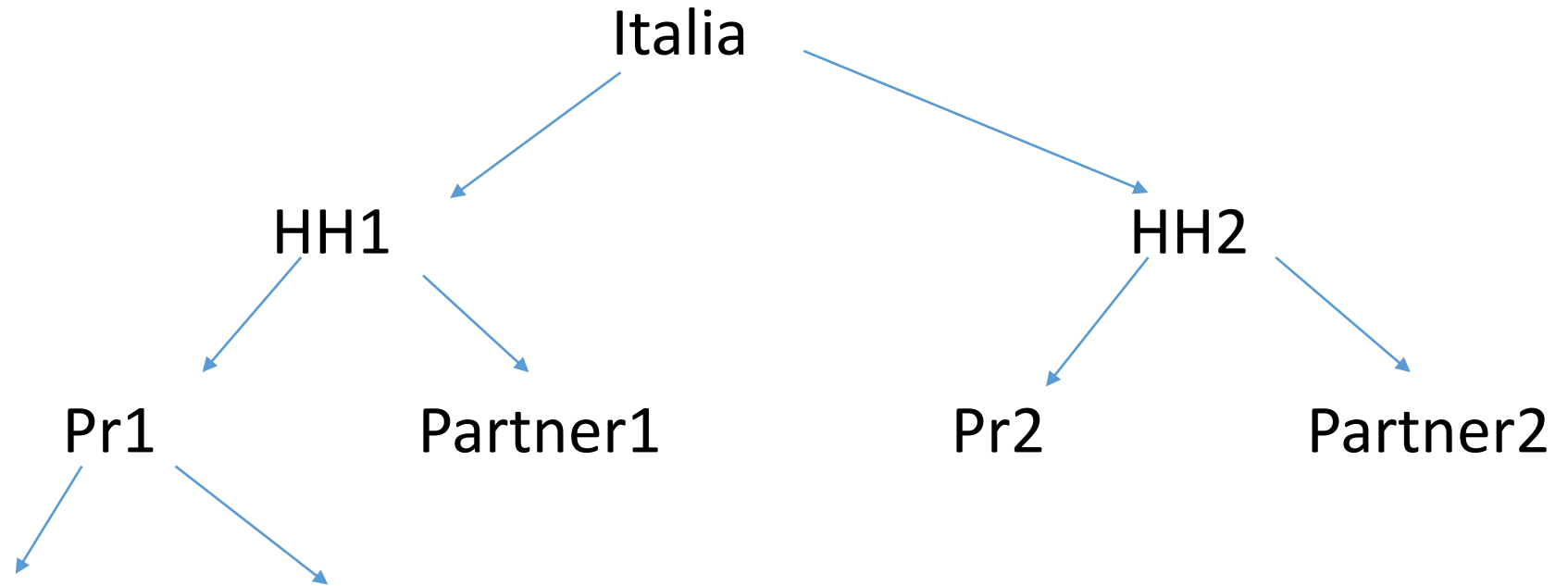
Struttura gerarchica

- Paese:

- Famiglie:

- Individui:

- Oss.: Pr1_2004 Pr1_2007....



Nel modulo CH gli intervistati rispondono a domande (es. dove abitano, lavoro, ecc...) sui figli (massimo 4 figli): è possibile costruire **Diadi genitori-figli**, aggiungendo quindi un altro livello (ogni Pr e Partner ha N diadi).

DATI – Easy SHARE

1 famiglia,
2 individui
(coppia)
4 osservazioni
(2004/2007)
ossia 2 per
individuo.

Re-sampling

Panel non-
bilanciato:
diverso
numero di
osservazioni
per individuo

	mergeid	wave	hhid	int_year	country
1	AT-000327-01	1	AT-000327-A	2004	11. Austria
2	AT-000327-01	2	AT-000327-A	2007	11. Austria
3	AT-000327-02	1	AT-000327-A	2004	11. Austria
4	AT-000327-02	2	AT-000327-A	2007	11. Austria
5	AT-000674-01	4	AT-000674-A	2011	11. Austria
6	AT-000674-01	5	AT-000674-A	2013	11. Austria
7	AT-000674-01	6	AT-000674-A	2015	11. Austria
8	AT-001215-01	4	AT-001215-A	2011	11. Austria
9	AT-001215-01	5	AT-001215-A	2013	11. Austria
10	AT-001215-01	6	AT-001215-A	2015	11. Austria
11	AT-001215-01	7	AT-001215-A	2017	11. Austria
12	AT-001492-01	4	AT-001492-A	2011	11. Austria
13	AT-001492-01	5	AT-001492-A	2013	11. Austria
14	AT-001492-01	6	AT-001492-A	2015	11. Austria
15	AT-001492-01	7	AT-001492-A	2017	11. Austria
16	AT-001492-02	4	AT-001492-A	2011	11. Austria
17	AT-001492-02	5	AT-001492-A	2013	11. Austria
18	AT-001492-02	6	AT-001492-A	2015	11. Austria
19	AT-001492-02	7	AT-001492-A	2017	11. Austria
20	AT-001816-01	1	AT-001816-A	2004	11. Austria
21	AT-001816-02	1	AT-001816-A	2004	11. Austria
22	AT-001816-02	2	AT-001816-A	2007	11. Austria
23	AT-001881-01	4	AT-001881-A	2011	11. Austria
24	AT-001881-01	5	AT-001881-A	2013	11. Austria
25	AT-001881-01	6	AT-001881-A	2015	11. Austria

A seconda delle analisi: 2 formati dei dati

- LONG e WIDE FORMAT

Figure 6.1: long vs. wide data format

Long format

mergeid	wave	int_year
AT-986403-01	1	2004
AT-986403-01	2	2007

Wide format

mergeid	int_year1	int_year2
AT-986403-01	2004	2007

INTRODUZIONE A STATA 13

Sintassi di base (linguaggio case-sensitive)

- help = spiegazione dei comandi + Manuale
- use/import/insheet = aprire base dati (save = sovrascrivere dati)
- reshape = cambiare formato dati (wide/ long)
- gen / egen = generare nuove variabili
- recode / replace = ricodificare / sovrascrivere variabili
- describe/sum / tab / tabstat/ fre = descrivere variabili
- hist/ bar / scatter = grafici
- decode/ encode = ricodificare variabili stringhe
- keep / drop = tenere/eliminare casi o variabili
- findit = trovare e installare comandi scritti da utenti
- reg/posson/logit/probit = modelli di regressione

Help e Manuale

Title

[D] `generate` — Create or change contents of variable

Syntax

Create new variable

```
generate [type] newvar[:lblname] =exp [if] [in]
```

Replace contents of existing variable

```
replace oldvar =exp [if] [in] [, nopromote]
```

Specify default storage type assigned to new variables

```
set type {float|double} [, permanently]
```

where `type` is one of

```
byte|int|long|float|double|str|str1|str2|...|str2045
```

See Description below for an explanation of `str`. For the other types,

`by` is allowed with `generate` and `replace`; see [D] `by`.

Menu

`generate`

Title

`generate` — Create or change contents of variable

Syntax

Remarks and examples

Menu

Methods and formulas

Description

References

Options

Also see

Syntax

Create new variable

```
generate [type] newvar[:lblname] =exp [if] [in]
```

Replace contents of existing variable

```
replace oldvar =exp [if] [in] [, nopromote]
```

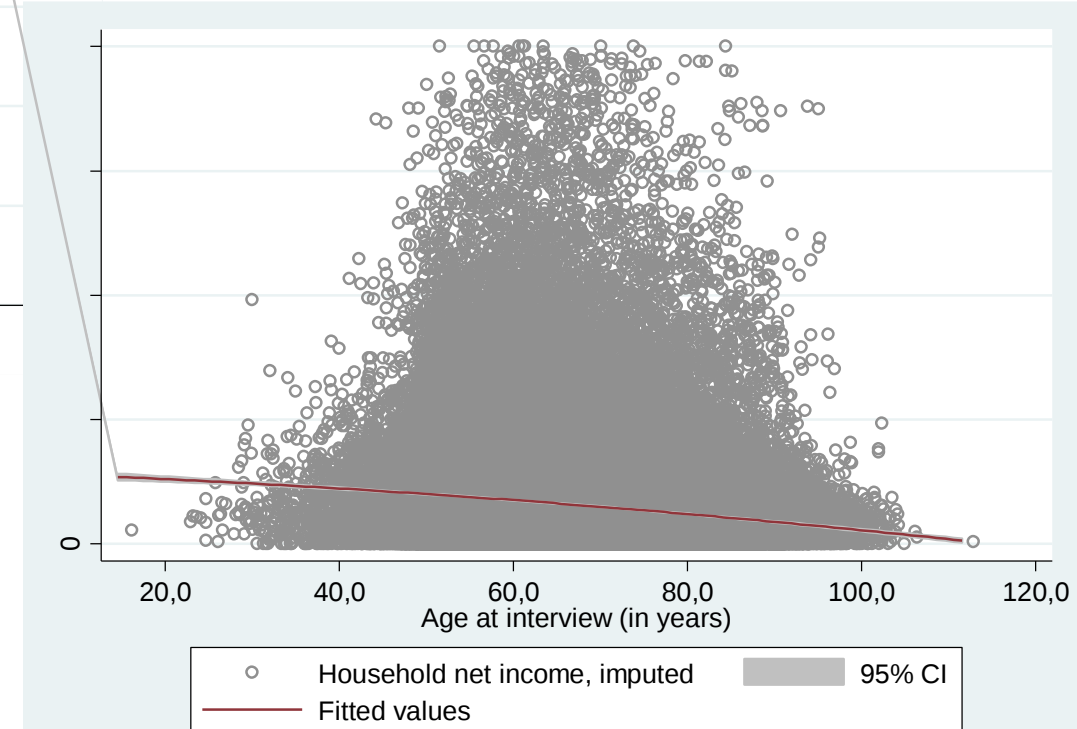
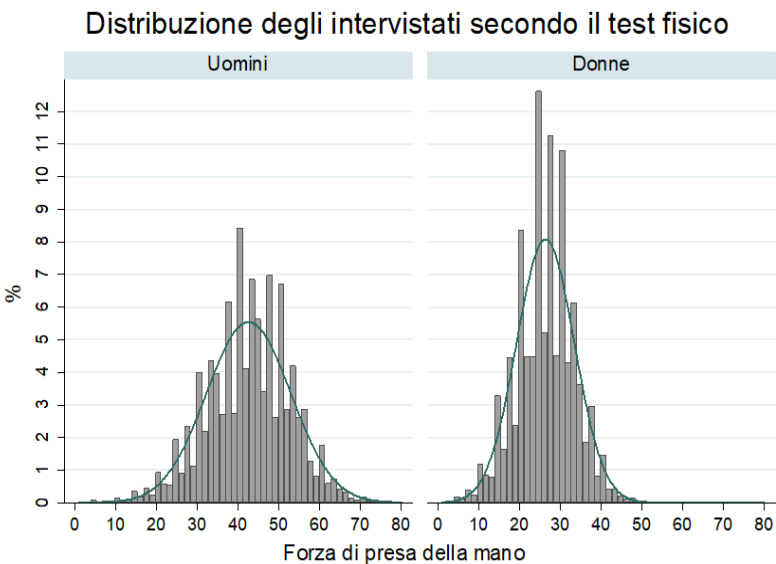
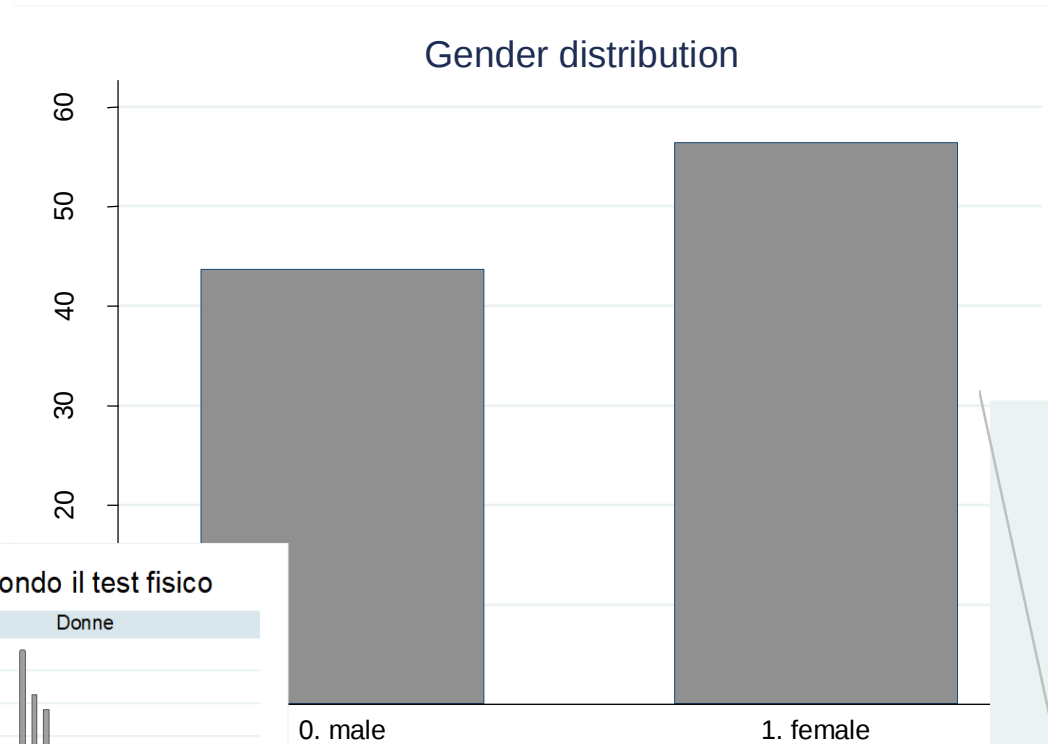
Specify default storage type assigned to new variables

```
set type { float|double } [, permanently]
```

where `type` is one of `byte | int | long | float | double | str | str1 | str2 | ... | str2045`.

See Description below for an explanation of `str`. For the other types, see [U] 12 Data.

Hist/bar/scatter



PURCHASE

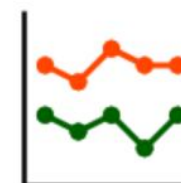
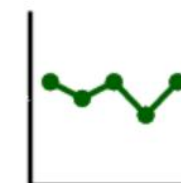
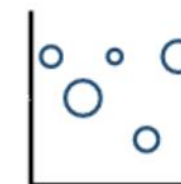
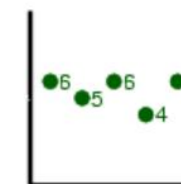
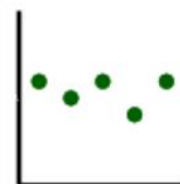
SUPPORT

COMPANY



Visual overview for creating graphs

Scatter and line plots



Risorse online (2)



About Statalist

We discuss Stata, statistics, and Stata and statistics.

You can [browse](#) without registering but you need to register to participate in the discussion or ask a question.

We encourage you to [register](#). We ask that you use your real name. Statalist has a long history of high quality discussion dating back to 1994. New-to-Stata members are welcome.

If you post a question, it will be more likely to be answered if you include the

How to participate

- [Register](#).
- [View without registering](#), which means you can browse without participating. You must register to post.
- [Enter the forum and then sign in](#) if you are already registered. Click **Login** at the top-right of the forum page.

Statalist is run and moderated by Stata users and maintained by StataCorp. If you have private comments about contents, policy, or the software, [contact the forum administrators](#).

Risorse online (3)



Institute for Digital Research & Education
Statistical Consulting

Search this website

HOME

SOFTWARE ▼

RESOURCES ▼

SERVICES ▼

ABOUT US

GRAPHING MEANS AND CONFIDENCE INTERVALS BY MULTIPLE GROUP VARIABLES | STATA CODE FRAGMENTS

The code below shows how to plot the means and confidence interval bars for groups defined by two categorical variables. This type of plot appeared in an article by Baker, et al, in *The American Journal of Clinical Nutrition*, "High prepregnant body mass index is associated with early termination of full and any breastfeeding in Danish women".

We start by generating the "bare bones" version of such a graph.

```
use https://stats.idre.ucla.edu/stat/stata/notes/hsb2, clear
```