

# SHARE data analysis - Italy

Sula Disem

## Contents

<b>1</b>	<b>Description of the Dataset for Italy</b>	<b>2</b>
<b>2</b>	<b>Measures of inequality and poverty based on household income</b>	<b>3</b>
<b>3</b>	<b>Measuring income inequality in Italy</b>	<b>4</b>
3.1	Gini Coefficient . . . . .	4
3.2	Theil index . . . . .	5
3.3	Palma ratio . . . . .	5
<b>4</b>	<b>Measuring poverty in Italy</b>	<b>6</b>
4.1	Headcount Ratio (poverty rate) . . . . .	6
4.2	Poverty Gap Index . . . . .	7
4.3	Poverty in subjective terms . . . . .	7
<b>5</b>	<b>Measuring inequality based on health outcomes</b>	<b>8</b>
5.1	Measures of health inequality . . . . .	8
5.1.1	Health disparity index . . . . .	9
5.2	Econometric analysis . . . . .	10
<b>6</b>	<b>Annex</b>	<b>13</b>
6.1	Link to the github repository of this project . . . . .	13
6.2	Source of the data . . . . .	13
6.3	References . . . . .	13

# 1 Description of the Dataset for Italy

The SHARE survey, or Survey of Health, Ageing and Retirement in Europe, is a multidisciplinary and cross-national research initiative aimed at understanding the dynamics of health, aging, and retirement among older adults in Europe and beyond. It aims to provide insights into the health, social, economic, and environmental factors influencing individuals' well-being and quality of life as they age. To do so it collects data from the same individual over multiple waves at regular intervals, to track changes in various aspect of life over time. The data is collected from a wide range of European countries, to enable comparison with trends and patterns across different cultures and economic contexts.

SHARE employs a rigorous **sampling design** to ensure representativeness of the older population in each participating country. The survey includes individuals from diverse socio-economic backgrounds and geographic regions.

After some manipulation of the data, filtering it to use only “Wave 7” and to consider only the country of choice “Italy”, the Dataset we will be using has **4568 rows** and **107 columns (variables)**

The following table provides some descriptive statistics of Income, Age and Years of education.

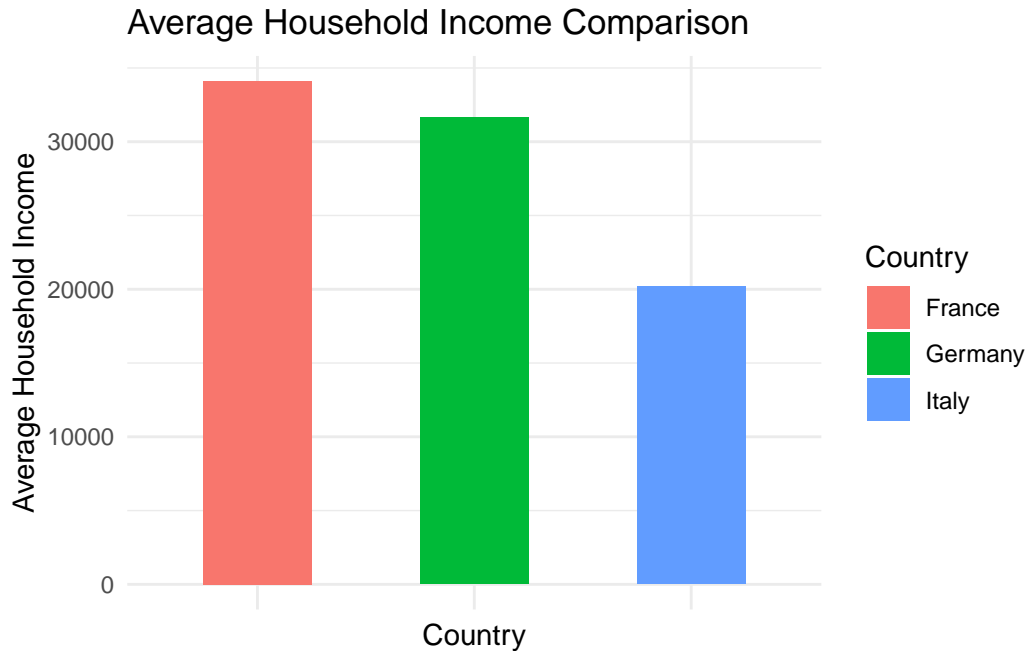
Variable	Min	X1st.Qu.	Median	Mean	X3rd.Qu.	Max
Income	0.0	12240.41	18149.62	20196.210000	25790.48	135517.2
Age	37.4	61.20	68.60	68.867251	76.20	97.8
Years of education	0.0	5.00	8.00	9.049281	13.00	25.0

Out of 4568 observations, we have here a table showing how the individuals are distributed among different types of employment status and gender:

	Count
Employed or Self-employed	928
Homemaker	850
Other	99
Permanently Sick or Disabled	112
Retired	2459
Unemployed	107
Female	2519
Male	2049

## 2 Measures of inequality and poverty based on household income

Let's now look into the average household income of Italy for wave 7 and compare it with two other countries: France and Germany.



The average household income for Wave 7 in Italy, France, and Germany are graphically compared above.

Based on the comparison, **France** has the highest average income ( $3.410713 \times 10^4$ ), followed not so far by **Germany** ( $3.1639452 \times 10^4$ ) and way lower by **Italy** ( $2.019621 \times 10^4$ )

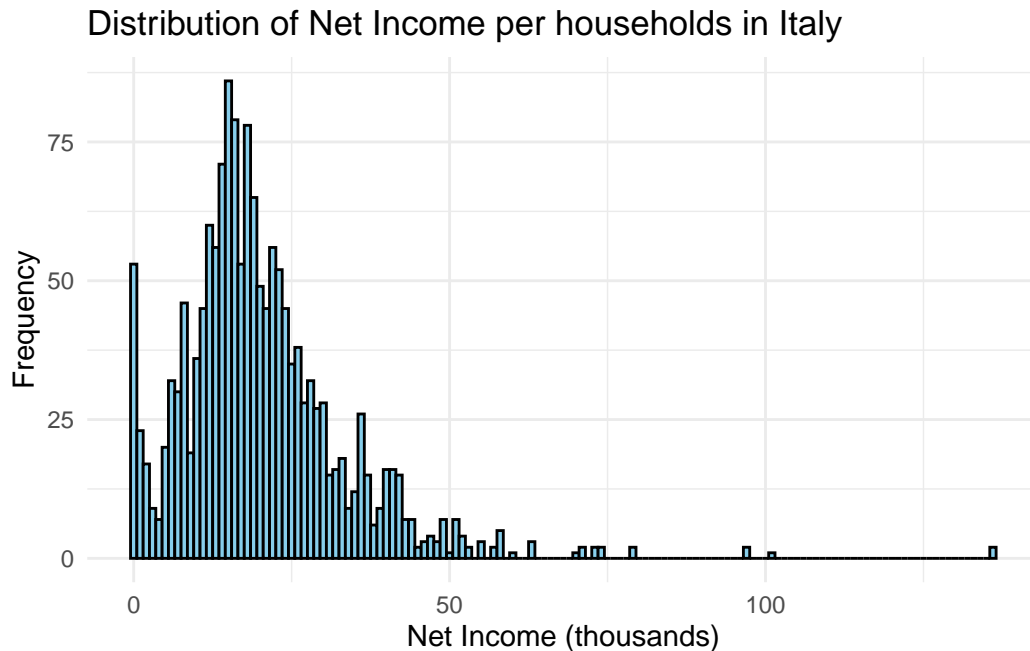
In the SHARE release guide for wave 7, some issues in terms of the imputations of wave 7 for household income have been pointed out. To compute the total household income for household that had mixed interviews, respondents to the other interview of “SHARELIFE” were treated as “nonresponding partners” (NRPs) in the regular interview. The problem is that detailed questions on current income were asked only to those with regular interview, so like standard NRPs issues, “incomes of all household members who answered the SHARELIFE interview are unobservable and focusing attention on the observable incomes of the regular respondents may introduce a severe downward bias in the aggregated measure of total household income.

In general, measuring household income is always difficult, because of misreporting or non response as cited before.

These issues could introduce bias and inaccuracies in household income estimates, leading to potentially unreliable rankings.

### 3 Measuring income inequality in Italy

Let's first have a look into the distribution of income in Italy with the help of a graph to have a general view of it, after that we will get into the measuring of inequality in net income of households in Italy, using Gini, Theil and Palma ratio.



#### 3.1 Gini Coefficient

To measure inequality let's start by using the **Gini** coefficient, an index to measure how much the distribution of income between individuals deviates from a perfectly equal distribution (Worldbank, Metadata Glossary). It is an index, so it provides a result between 0 and 1, a higher Gini coefficient indicates higher income inequality, while a coefficient closer to 0 indicates lower income inequality.

```
[1] "Gini coefficient for Italy: 0.33"
```

We can see that for our country of references we find a gini coefficient equal to 0.3339782 (0.334), which indicates some inequality but not that much. The Gini coefficient is the most widely used measure of income inequality, because it is very easy to calculate and interpret. The problem is that it does not provide any insights into the distribution of income at different parts of the income distribution

### 3.2 Theil index

We are now going to consider another coefficient to measure inequality: **The Theil Index**. The Theil index measures the distance between what we observe in the data and a situation where everyone has the same income. The closer we get to 0 the more equality, the closer to 1, the higher the level of inequality.

```
[1] "The Theil index for Italy: 0.18"
```

We see a Theil index that again suggests not that high inequality of incomes in Italy. The problem with this index is that it is sensitive to the size of the population and the weights you put in each group, even though in this case we didn't assign any way.

### 3.3 Palma ratio

A last measure of income inequality is a share ratio, called "Palma ratio", it looks at where most of the inequality comes from, does it come from the fact that the top 10% have a way higher income, and/or from the fact that the bottom 40% have a very low income?

```
[1] "The Palma ratio 90/40 for Italy: 1.26"
```

This result for Italy, a Palma ratio of approximately 1.26, means that the income share of the top 10% of the population is about 1.26 times larger than the income share of the bottom 40%. It suggests that a significant portion of the total income is concentrated among the wealthiest individuals, something that is not surprising in developed countries like Italy.

The Palma ratio is very useful when looking into extreme income shares, to really understand where most of the inequality comes from. But it has to be used carefully, as it may oversimplify the income distribution by focusing on only two segments of the population (in our case, top 10% and bottom 40%).

## 4 Measuring poverty in Italy

Let's now focus on the bottom of the distribution by looking into different measures of poverty. Even though is not that easy to measure it, especially when having to decide what is the threshold in income to be considered in poverty for example. We can define it in monetary terms by looking at it in **absolute** terms (lack of minimum resources like food), **relative** (to living standards) and **subjective terms** ("do you think you have enough to meet ends?") We first need to find out poverty line, which in Italy changes accordingly to the fact that you live in the north, centre or south and of course it changes accordingly to the number of components in the family and their age. Since we do not know from where the individual come from, how many components in a household, the best thing to do is find a **relative poverty line** calculated as a percentage of the median income. It could be interesting in this case to create a relative poverty line at 60% of the median net income of the household of the sample in the data we are using.

```
[1] "Relative Poverty Line: 10889.77"
```

This poverty line is quite consistent with the poverty line stated by ISTAT in Italy, considered equal to the average monthly pro-capite expenditure in 2017 for a household composed by one component (Instat, 2018).

### 4.1 Headcount Ratio (poverty rate)

With the poverty line previously determined, we can now look at the poverty rate, calculated as:

$$\text{Total'under'the'poverty'line/totalpopulation}$$

This is what we get:

```
[1] "Headcount Ratio (Poverty Rate): 19.39 %"
```

This is a quite simple and easy to understand measure, but its limitations are obvious. First it does not say anything about the depth of poverty, you are either in poverty or no depending on where the poverty line has been fixed. As we previously said, where the individual are living in Italy changes the poverty line by quite a lot.

## 4.2 Poverty Gap Index

Let's now have a look into where the poor stand in terms of distance from the poverty line, by using the poverty gap index. This index measures the depth of poverty by considering the distance separating the income of the poor from the poverty line.

It is calculated as:

$$PGI = 1/N \sum (povertyline - income_{of\ poor}) / povertyline$$

```
[1] "Poverty Gap Index: 0.1"
```

It shows that on average, incomes of individuals below the poverty line are 10% below that threshold. Of course with this result only we can't say anything about the inequality inside the population under the poverty line, It might also be a problem the fact that we have many zeros in the income stated by some individuals, which would drag down some calculation and might bias our measures, because those zeros are probably just non replies.

## 4.3 Poverty in subjective terms

We have measured poverty in relative terms until now, by selecting a threshold to state if a household has to be considered in poverty or no. Now we are going to look into poverty in more **subjective terms**, by looking into the households ability to make ends meet. Individuals in the survey have been asked: "Thinking of your household's total monthly income, would you say that your household is able to make ends meet...". Each of them could answer: 1 (great difficulty), 2 (some difficulty), 3 (fairly easily), 4 (easily). We are going to consider individuals answering 1 and 2 as in experiencing some level of financial issues, which could be indicative of poverty, and we are going to compute a subjective poverty rate.

```
[1] "Subjective Poverty Rate: 52.58 %"
```

We find that 52.58% of individuals in the survey state that their household is able to make ends meet "with great difficulty" or "some difficulty". Since individuals might overestimate their issue when saying "some difficulty", let's consider in poverty only individuals that make ends meet "with great difficulty":

```
[1] "Subjective Poverty Rate: 17.08 %"
```

This result is much more consistent with the previous relative poverty measures that we have found so far.

## 5 Measuring inequality based on health outcomes

We are now going to measure health inequality, but let's first have a look at the self-perceived health, where respondents had to answer about their health status with: 1) Excellent, 2) Very good, 3) Good, 4) Fair, 5) Poor

	Satisfaction_Level	Frequency	Percentage
1	1	332	7.282299
2	2	689	15.112963
3	3	1537	33.713534
4	4	1548	33.954815
5	5	453	9.936390

We can see that most of the respondents answer between fair and good when asked about their self-reported health, while most of them refrain from answering “excellent” to that question.

Before getting into the measure of health inequality, let's look into a variable that is less subjective than the previous one. `chronic_mod` is a variable that looks into the number of chronic diseases of individuals. Respondents had to answer to the question: “Has a doctor ever told you that you had ...”, considering the following chronic diseases: 1. A heart attack; 2. High blood pressure or hypertension; 3. High blood cholesterol; 4. A stroke or cerebral vascular disease; 5. Diabetes or high blood sugar; 6. Chronic lung disease (...); 10. Cancer or malignant tumor (...); 11. Stomach or duodenal ulcer, peptic ulcer; 12. Parkinson disease; 13. Cataracts; 14. Hip fracture or femoral fracture. Individuals that did not have any of those have 0 as result, so we can look into the percentage of respondents that has had a chronic disease and that has not:

The data shows that more than 60% of respondents for Italy have had at least one of the chronic diseases mentioned, it looks quite consistent with the low level of self-reported health we saw before.

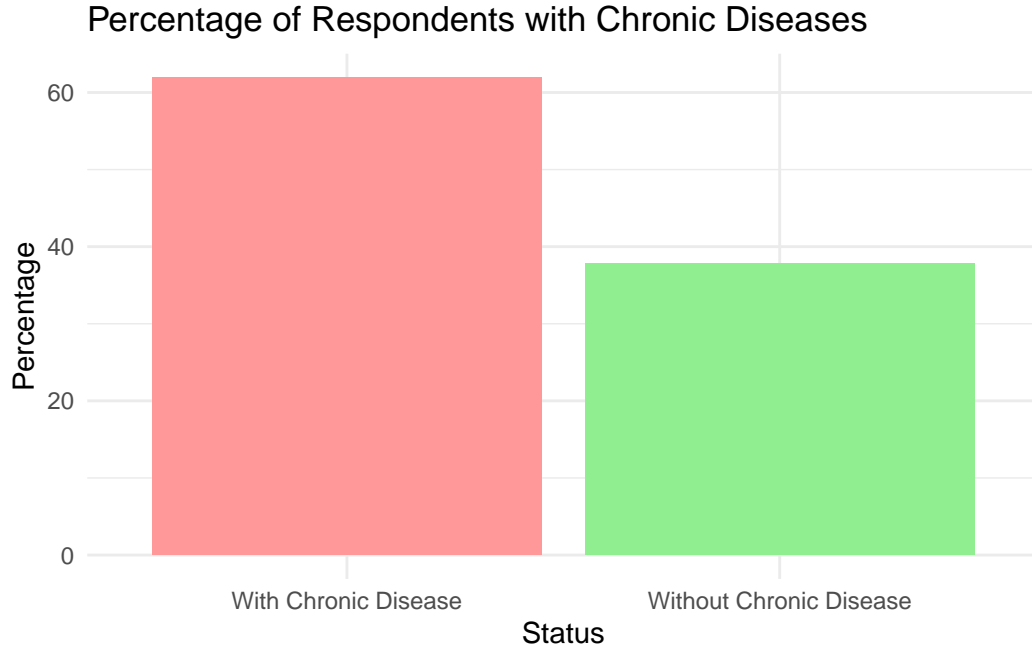
With this general view of our country of reference, we are now going to measure health inequality at the individual level.

### 5.1 Measures of health inequality

We first have to select some health indicators, from the data available we can consider:

- *Self-perceived health status*
- *Presence of chronic diseases*
- *Limitations in activities of daily living (ADL)*, which indicates difficulties in performing basic self-care tasks, from 0 to 5, the higher the index the higher the difficulty of the respondent to complete these tasks.





- *Limitations in instrumental activities of daily living (IADL)*, which indicates the difficulty of performing more difficult activities necessary for independent living, ranges from 0 to 5, the higher the more difficulties in doing these tasks.
- *Depression scale*, which measures the current depression, based on several measures: depressed mood, pessimism, suicidality, guilt, sleep, interest, irritability, appetite, fatigue, concentration, enjoyment and tearfulness. It ranges from 0 (not depressed), to 12 (very depressed)

We find some subgroups, in our case what would be interesting to see is in terms of education, sex and place of residence (City, village, ...)

The health inequality measures that we will use are reported by the review of existing measures by the World Helth Organization (Schlotheuber and Ahmad Reza Hosseinpoor, 2022).

### 5.1.1 Health disparity index

We will use two different indexes, first the health disparity index, a relative measure of inequality that shows the average difference between each population subgroup and the setting average, in relative terms.

Gender	Self_Reported_Health	ADL	Chronic_Disease
Male/Female	0.98	0.69	1.13

The health disparity per gender shows that there is almost no inequality in the self reported health between men and women, while we see that man are more likely to have at least one chronic disease.

We do the same for the place of residence and here are the results:

Area	Self_Reported_Health	ADL	Chronic_Disease
City/Suburbs	0.97	0.56	0.96
City/Large Town	0.94	0.61	0.87
City/Small Town	0.94	0.60	0.93
City/Rural Area	0.92	0.51	0.76

The first thing we notice is that again we don't find much of inequality in self reported health between respondents resident in different areas. It seems to be high inequality in terms of difficulties of carrying activities of daily living between people living in the city and those living outside. For what it concerns chronic disease, it seems that individuals living in rural areas suffer more from these diseases, even though the difference is not as high as for ADL.

## 5.2 Econometric analysis

Health inequalities across individuals can be influenced by various socioeconomic, demographic, and environmental factors. To conduct an econometric analysis, we can employ a regression model to explore the relationship between these factors and health outcomes.

We will be looking at the self reported health and how this variable is affect by several specifications, **demographic** (such as age and gender), **socioeconomic** (income, education, occupation), and health related (diet, phisical activity, smoking).

The regression model used will be an ordered probit model, given the characteristics of the dependent variable, which takes 5 outcomes that can be ordered from low to high self reported health.

The econometric model writes as follow:

---


$$Health_i = \beta_0 + \beta_1 \times Income_i + \beta_2 \times Education_i + \beta_3 \times Age_i + \beta_4 \times bmi_i + \beta_5 \times Smoking_i + \beta_6 \times Gender_i + \epsilon_i$$


---

Figure 1: Econometric Model Equation

Where:

- ( Health\_i ) is the health outcome for individual ordered from 1 (exellent) to 5 (poor).
- ( Income\_i ) is the income level for individual that we will separate in 5 subgroups.

- ( Education\_i ) is the education level for individual (in years of study).
- ( Age\_i ) is the age of individual.
- ( bmi\_i ) is the bmi of individual.
- ( Smoking\_i ) whether the individual smokes or not.
- ( Gender\_i ) is the gender of individual.
- ( beta\_0, beta\_1, beta\_2, beta\_3, beta\_4 ) are the coefficients to be estimated.
- ( \_i ) is the error term.

This model allows us to estimate the relationship between socioeconomic and demographic factors and health outcomes. With a probit model we can not look into the size of the effects but we can say if an independent variable increases or decreases the likelihood for the individual to report high or low level of self reported health. To do so we have to create brackets for the variable income, we will use the progressive tax rate “IRPEF” of Italy, we will have so 5 brackets, from low income to very high.

[1] TRUE

Call:

```
polr(formula = sphus ~ income_bracket + edueyears_mod + age +
      bmi + smoking + female, data = New_Italy_omitted, method = "probit")
```

Coefficients:

	Value	Std. Error	t value
income_bracket	-0.08144	0.047305	-1.7216
edueyears_mod	-0.02320	0.009519	-2.4374
age	0.02583	0.004583	5.6369
bmi	0.01955	0.007757	2.5207
smoking	0.03890	0.027528	1.4133
female	0.06167	0.073832	0.8353

Intercepts:

	Value	Std. Error	t value
1 2	0.5679	0.4517	1.2572
2 3	1.1313	0.4514	2.5061
3 4	2.2050	0.4539	4.8584
4 5	3.4945	0.4588	7.6173

Residual Deviance: 2651.065

AIC: 2671.065

(519 observations deleted due to missingness)

Keeping in mind that for self perceived health respondents answered with a number between 1 to 5, with 1 being “Excellent” and 5 being “Poor”. It seems that income and education negatively impact the probability of perceiving lower levels of health, while bmi and age increases the likelihood of perceiving a lower level of health. Nothing clear can be said about gender and whether the respondent is a smoker or not.

The data unfortunately is made of a lot of missing values, we have to omit them to be able to use our model. Ignoring missing values can lead to biased estimates, distorting results and conclusions. Additionally it reduced the sample size, potentially removing important information and providing less strong results.

## 6 Annex

### 6.1 Link to the github repository of this project

[Github repository](#)

### 6.2 Source of the data

This paper uses data from the generated easySHARE data set (DOI: [10.6103/SHARE.easy.800](#)), see Gruber et al. (2014) for methodological details. The easySHARE release 8.0.0 is based on SHARE Waves 1, 2, 3 (SHARELIFE), 4, 5, 6, 7 and 8 (DOIs: [10.6103/SHARE.w1.800](#), [10.6103/SHARE.w2.800](#), [10.6103/SHARE.w3.800](#), [10.6103/SHARE.w4.800](#), [10.6103/SHARE.w5.800](#), [10.6103/SHARE.w6.800](#), [10.6103/SHARE.w7.800](#), [10.6103/SHARE.w8.800](#))

### 6.3 References

- Börsch-Supan, A. & S. Gruber (2020): *easySHARE. Release version: 8.0.0. SHARE-ERIC. Dataset.* doi: [10.6103/SHARE.easy.800](#)
- Istat 2018, *Condizione di vita, reddito e carico fiscale delle famiglie*
- Asada, Y., Hurley, J., Norheim, O.F. et al. *A three-stage approach to measuring health inequalities and inequities.* Int J Equity Health 13, 98 (2014). <https://doi.org/10.1186/s12939-014-0098-y>