



Clustering expenditure patterns: an application to the Portuguese Household Budget Survey

Eduarda Góis, Maria Manuel Pinho, Cristina Gonçalves, Carla Afonso

Statistics Portugal

Living Conditions Statistics Unit

JOCLAD 2025

Porto | April 4, 2025

Outline

- 1. The Portuguese Household Budget Survey (HBS)**
- 2. The 2022/2023 HBS main results**
- 3. A cluster analysis on household expenditure patterns**

1. The Portuguese HBS

- The national HBS is part of the European HBS project and is in line with the **recommendations agreed between Eurostat and the Member States**
- The main goal is the five-year calculation of the structure of **household expenditure on goods and services according to COICOP** (Classification of Individual Consumption by Objective), contributing to the:
 - updating of the weights of the **Consumer Price Index**
 - estimation of the **National Accounts** private consumption
 - approximating the **population diet** using the quantities of food purchased

1. The Portuguese HBS

- It was launched at the national level in the late 1960s, and since then there have been **10 rounds**
 - one of Statistics Portugal's most **consolidated large-scale statistical survey**, targeting a representative sample of the population living in the national territory and stratified regionally (NUTS 2 level)
- In addition to the survey results available at Statistics Portugal and open to the public, the **anonymised data files, available for research purposes**, can be used for detailed analyses on the allocation of Portuguese households' expenditure

www.ine.pt | Products | Microdata

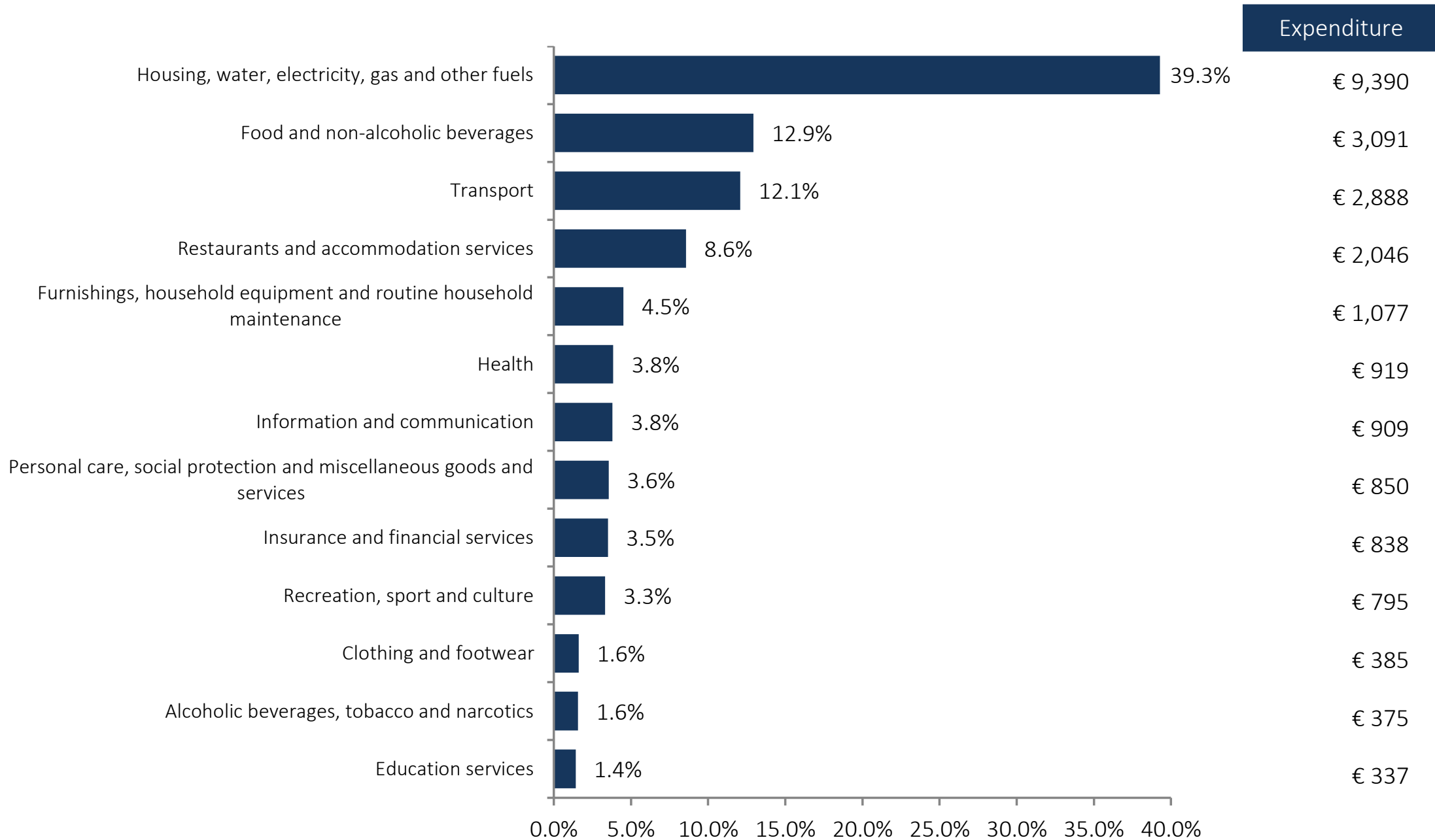
1. The Portuguese HBS

- For future editions, Statistics Portugal intends to make use of the **increasing availability of administrative data**
- Our aim is to explore the data from the Tax Authority
 - specifically **electronic invoices** asked by final consumers on household expenditures (collected through the e-fatura system)
 - and information on household expenditure from **tax deductions** and **electronic invoices on housing rents**
 - to validate and eventually calibrate household expenditure data
- The main advantages are the mitigation of the **respondent burden**, lower associated **costs**, and the availability of data for the entire **statistical population**

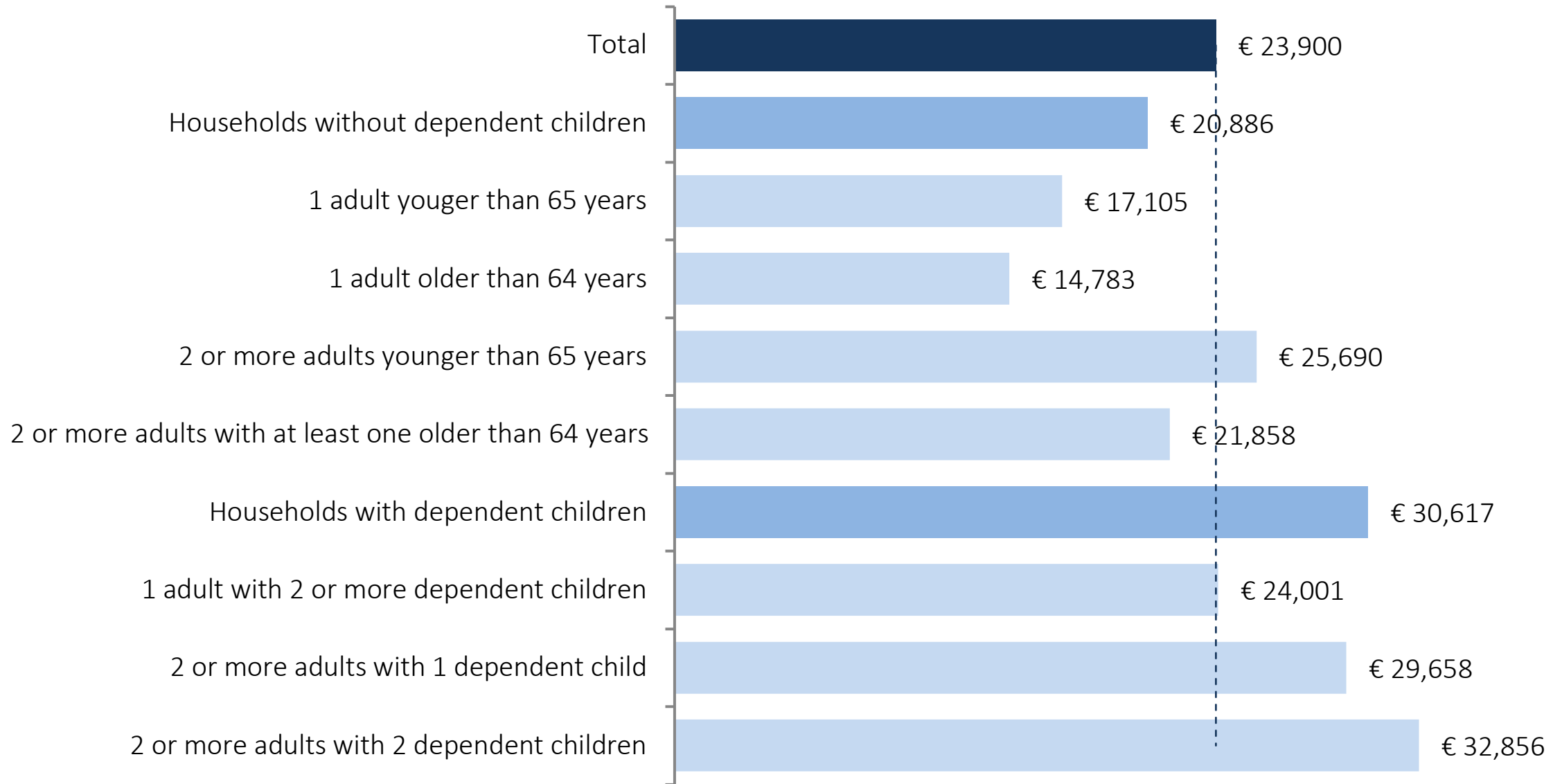
2. The 2022/2023 HBS main results

- Mean annual household expenditure: **€23,900**
- About 2/3 of which were concentrated in **housing** (39.3%), **food** (12.9%) and **transport** (12.1%)
- **Households with children** spend annually on average €9,731 more than those without children (a mean monthly expenditure €811 higher)
- Households in the **top income quantile** (20% of households with the highest incomes) spent more than twice as much as households in the bottom income quantile (20% of households with the lower incomes)

Mean annual total expenditure structure by COICOP division, Portugal, 2022/2023



Mean annual total expenditure by household type, Portugal, 2022/2023



2. The 2022/2023 HBS main results

- The technical aspects underlying the 2022/2023 edition of the survey make it impossible to ensure consistency in HBS time series
- Even so, it was possible to carried out a **simplified exercise** using a correspondence between the two versions of the classification made available by Eurostat
- The most evident result is the **increase in importance of housing expenditures**: the share on total expenditure practically doubled between 2000 and 2022/2023
- In the opposite direction, **expenditure on food and on clothing and footwear has lost relevance** in the structure of household expenditure, which has been extended, albeit to a lesser extent, to expenditure on **transport**

Mean annual total expenditure structure by eCOICOP division, Portugal

eCOICOP	2000	2005/2006	2010/2011	2015/2016	2022/2023
01 - Food and non-alcoholic beverages	18,7%	15,5%	13,3%	14,3%	12,9%
02 - Alcoholic beverages, tobacco and narcotics	2,8%	2,3%	1,9%	1,6%	1,6%
03 - Clothing and footwear	6,6%	4,1%	3,7%	3,5%	1,6%
04 - Housing, water, electricity, gas and other fuels	19,8%	26,6%	29,2%	31,9%	39,2%
05 - Furnishings, household equipment and routine household maintenance	7,2%	4,8%	4,2%	4,0%	4,6%
06 - Health	5,2%	6,1%	5,8%	5,5%	3,9%
07 - Transport	15,0%	12,9%	14,5%	14,1%	12,1%
08 - Communication	3,3%	3,0%	3,3%	3,2%	3,3%
09 - Recreation and culture	4,8%	5,7%	5,3%	4,2%	3,8%
10 - Education	1,3%	1,7%	2,2%	2,3%	1,4%
11 - Restaurants and hotels	9,5%	10,8%	10,4%	8,8%	8,6%
12 - Miscellaneous goods and services	6,1%	6,5%	6,3%	6,7%	7,0%

3. A cluster analysis on household expenditure patterns

- The aim is to create **homogeneous groups of households according to their expenditure pattern**
 - internally cohesive and externally isolated groups
 - homogeneity within clusters and heterogeneity between clusters
- The **grouping variables** are the previously standardised **monetary expenditures** (in euros) for the **COICOP first level** (divisions), with the application of the survey weighting
- **Excluding non-monetary expenditure** from the grouping variables implies disregarding self-consumption, income in kind from employment and owner-occupied imputed housing rents

3. A cluster analysis on household expenditure patterns

As for the grouping variables

- we excluded **Division 13**, *Personal care, social protection and miscellaneous goods*, since it has a residual component and no discriminatory power
- we summed **Division 09**, *Recreation, sport and culture*, and **Division 11**, *Restaurants and accommodation services*, since they capture to some extent similar consumption purposes
- in line with the classification of housing acquisition as an investment expenditure, we also decided to remove from the grouping variables the purchase of vehicles (**Group 071**) and of recreative durable goods (**Class 0912**)

3. A cluster analysis on household expenditure patterns

- given the high number of 2022/2023 cases (**11,701 households in the sample**), a non-hierarchical classification technique is used, the **K-means** method, consisting of transferring an individual to the cluster whose centroid is at the shortest distance
 - for large datasets, hierarchical techniques are far more computationally demanding
- to achieve the best differentiation and interpretative interest regarding the expenditure profile of the households, our choice is to produce four clusters (**K=4**)

Cluster 1

Mean annual expenditure

total: €72,377

monetary: €48,758

Monetary expenditure

22.0% in education

14.3% in transport

11.9% in restaurants
and accommodation

10.0% in household
equipment and
maintenance

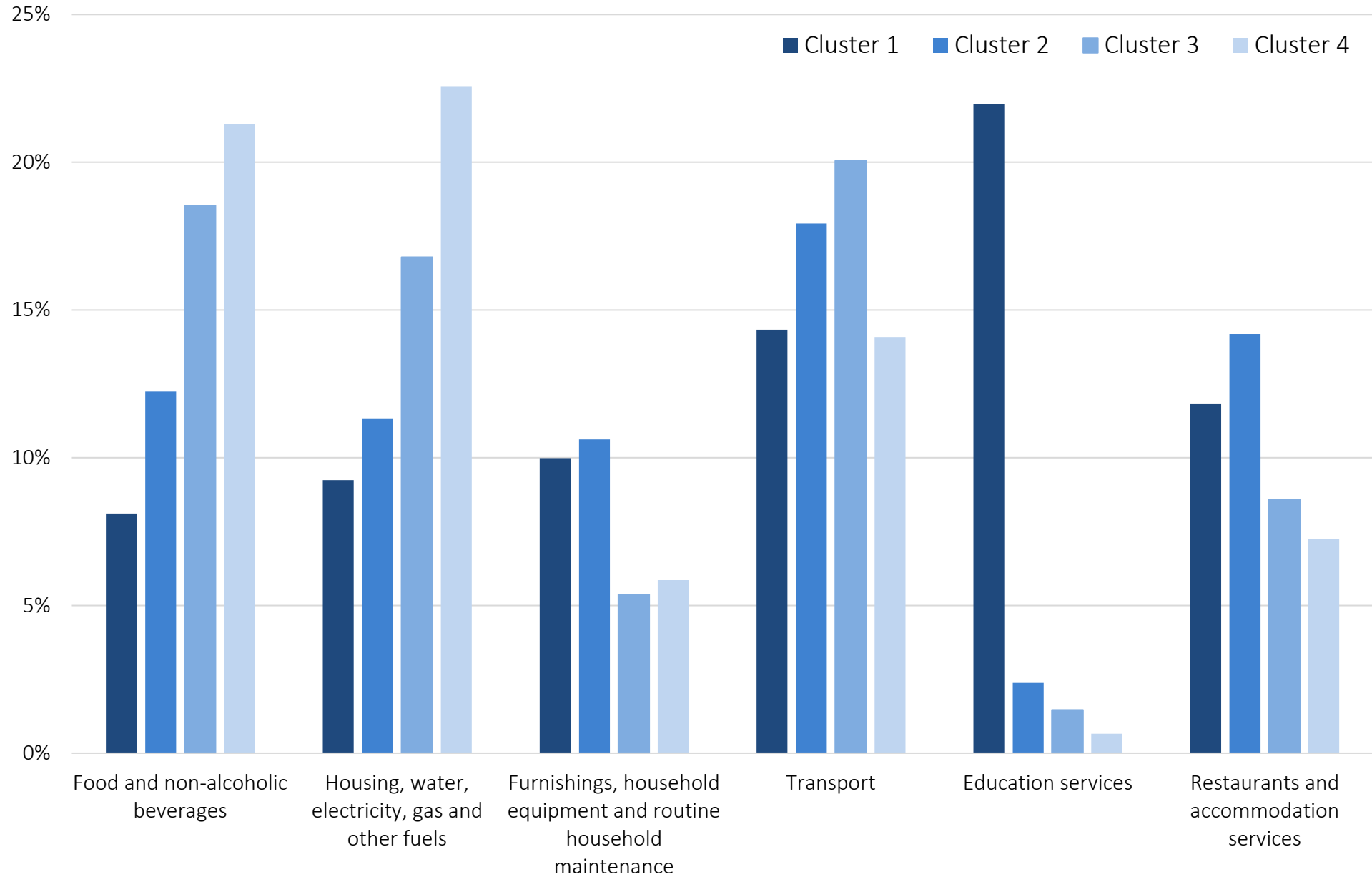
- represents 1% of the households
- average household size: 4.0 people
- 97% included children
- 73% in the 5th income quantile
- 96% owned their home

Cluster 1	Cluster 2
<p>Mean annual expenditure total: €72,377 monetary: €48,758</p>	<p>Mean annual expenditure total: €54,542 monetary: €39,710</p>
<p>Monetary expenditure 22.0% in education 14.3% in transport 11.9% in restaurants and accommodation 10.0% in household equipment and maintenance</p>	<p>Monetary expenditure 17.9% in transport 14.2% in restaurants and accommodation 12.2% in food 11.3% in housing 10.6% in household equipment and maintenance</p>
<ul style="list-style-type: none"> - represents 1% of the households - average household size: 4.0 people - 97% included children - 73% in the 5th income quantile - 96% owned their home 	<ul style="list-style-type: none"> - represents 7% of the households - average household size: 3.5 people - 59% included children - 61% in the 5th income quantile - 93% owned their home

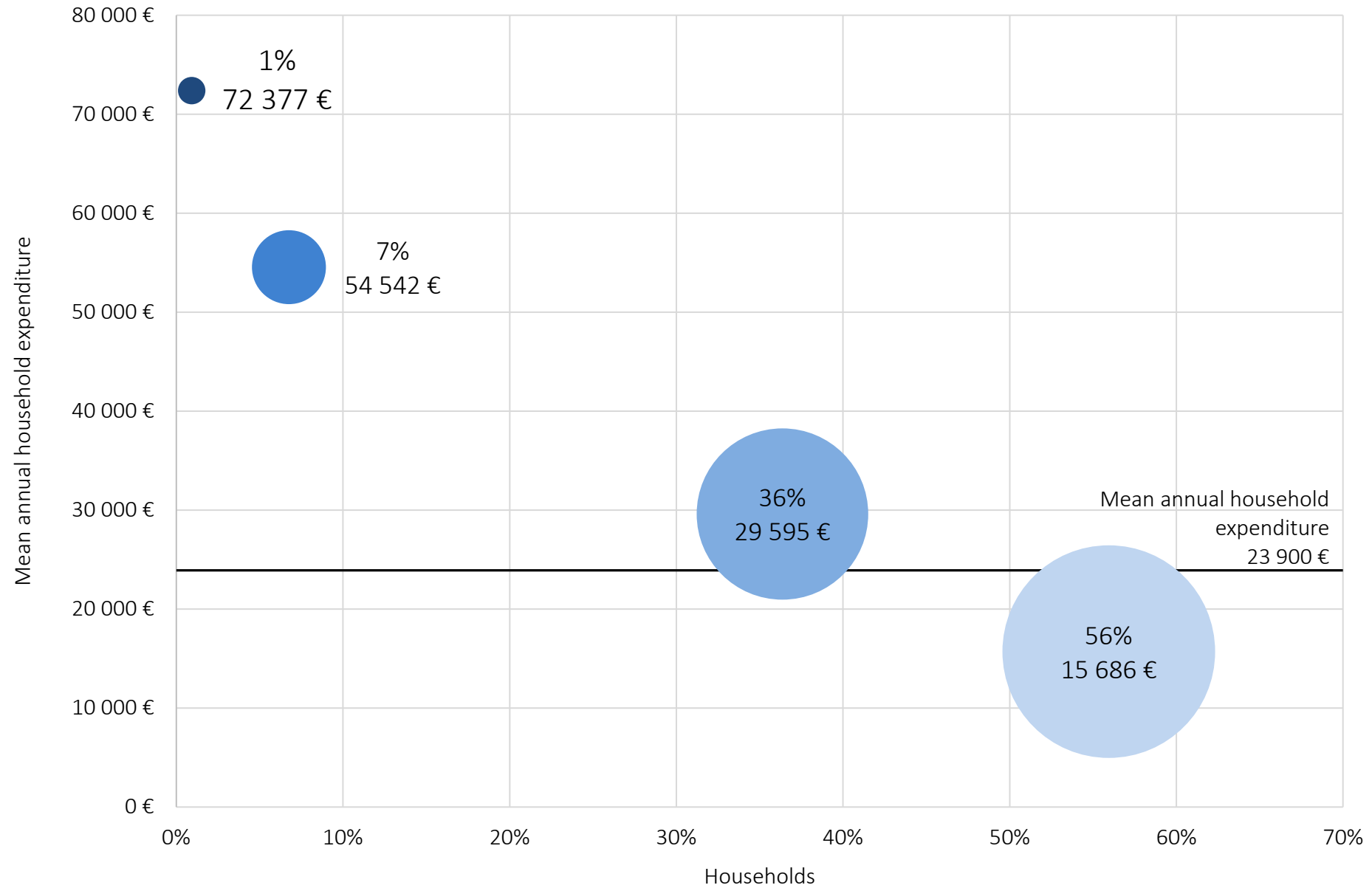
Cluster 1	Cluster 2	Cluster 3
<p>Mean annual expenditure total: €72,377 monetary: €48,758</p>	<p>Mean annual expenditure total: €54,542 monetary: €39,710</p>	<p>Mean annual expenditure total: €29,595 monetary: €21,162</p>
<p>Monetary expenditure 22.0% in education 14.3% in transport 11.9% in restaurants and accommodation 10.0% in household equipment and maintenance</p>	<p>Monetary expenditure 17.9% in transport 14.2% in restaurants and accommodation 12.2% in food 11.3% in housing 10.6% in household equipment and maintenance</p>	<p>Monetary expenditure 20.1% in transport 18.5% in food 16.8% in housing 8.6% in restaurants and accommodation</p>
<ul style="list-style-type: none"> – represents 1% of the households – average household size: 4.0 people – 97% included children – 73% in the 5th income quantile – 96% owned their home 	<ul style="list-style-type: none"> – represents 7% of the households – average household size: 3.5 people – 59% included children – 61% in the 5th income quantile – 93% owned their home 	<ul style="list-style-type: none"> – represents 36% of the households – average household size: 3.1 people – 45% included children – 53% in the 4th or 5th income quantile – 77% owned their home

Cluster 1	Cluster 2	Cluster 3	Cluster 4
<p>Mean annual expenditure total: €72,377 monetary: €48,758</p>	<p>Mean annual expenditure total: €54,542 monetary: €39,710</p>	<p>Mean annual expenditure total: €29,595 monetary: €21,162</p>	<p>Mean annual expenditure total: €15,686 monetary: €9,231</p>
<p>Monetary expenditure 22.0% in education 14.3% in transport 11.9% in restaurants and accommodation 10.0% in household equipment and maintenance</p>	<p>Monetary expenditure 17.9% in transport 14.2% in restaurants and accommodation 12.2% in food 11.3% in housing 10.6% in household equipment and maintenance</p>	<p>Monetary expenditure 20.1% in transport 18.5% in food 16.8% in housing 8.6% in restaurants and accommodation</p>	<p>Monetary expenditure 22.6% in housing 21.3% in food 14.1% in transport 7.2% in restaurants and accommodation</p>
<ul style="list-style-type: none"> - represents 1% of the households - average household size: 4.0 people - 97% included children - 73% in the 5th income quantile - 96% owned their home 	<ul style="list-style-type: none"> - represents 7% of the households - average household size: 3.5 people - 59% included children - 61% in the 5th income quantile - 93% owned their home 	<ul style="list-style-type: none"> - represents 36% of the households - average household size: 3.1 people - 45% included children - 53% in the 4th or 5th income quantile - 77% owned their home 	<ul style="list-style-type: none"> - represents 56% of the households - average household size: 2.0 people - 17% included children - 29% in the 4th or 5th income quantile - 73% owned their home

Mean annual total expenditure structure for some COICOP divisions per cluster



Mean annual total expenditure and share of households per cluster



Thank you!



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL