HOUSEHOLD FINANCIAL SURVEY 2024

banco central Chile CIEN años

Methodological Report

September 2025

Contents

Pág.

I. Executive Summary

The Household Financial Survey (HFS), conducted by the Central Bank of Chile since 2007, aims to generate detailed information on the financial balance of households and monitor its evolution over time. The analysis of these data enables a deeper understanding of household financial behavior, which is relevant to the Central Bank's objectives of financial and price stability.

This document provides a general overview of the main methodological aspects of the 2024 edition of the HFS, whose structure is outlined below.

First, the 2024 HFS questionnaire is presented, including the contents of each module that comprises the survey and the key definitions that determine the household and the respondent. It also describes the main modifications compared to the 2021 HFS questionnaire, which was adapted for telephone application during the pandemic.

Second, the sampling design implemented in the 2024 HFS is described. As a rotating panel survey, each round includes a panel sample (households interviewed in the previous round) and a refreshment sample (households interviewed for the first time in the current round). The sampling design of the refreshment sample is detailed, and the process of contacting the panel sample prior to the 2024 fieldwork is summarized.

Third, the main characteristics and results of the fieldwork are presented. The process concluded with 4,649 successful interviews: 1,801 from the panel sample and 2,848 from the refreshment sample.

Next, the calculation of expansion factors is explained, which are necessary for the statistics derived from the 4,649 interviews to be representative of urban households at the national level. The calculation is performed separately for each sample and then calibrated to the total population. To obtain robust estimates, the variance estimation process is detailed, along with the methodology used and its main results.

Finally, the imputation methodology applied, and its most relevant results are presented.

II. Introduction

Households are key actors in both the financial system and the broader economy. Their financial decisions are closely linked to other economic choices, such as those related to the labor market.

Analyzing household finances allows for a deeper understanding of their behavior within the financial system and supports the development of tools to monitor their financial situation. Both dimensions are relevant to the Central Bank's objectives of financial stability and price control.

In this context, since 2007 the Central Bank of Chile has conducted the Household Financial Survey (HFS), aimed at generating detailed information on household financial balance.

Between 2007 and 2024, nine rounds of the HFS have been conducted. The editions of 2007, 2011–2012, 2014, 2017, 2021, and 2024 are representative at the national urban level, while the 2008, 2009, and 2010 editions were limited to the urban area of the Metropolitan Region.

Like other financial surveys[1], the HFS has specific methodological requirements. First, it follows a rotating panel design, with a complex sampling structure that includes oversampling of high-wealth households. Second, it seeks to collect sensitive financial information from households. Lastly, the expected levels of non-response make planning and contact strategies critical to the quality of the results.

The sixth national edition of the HFS was conducted during the second half of 2024, achieving a total of 4,649 interviews with urban households across all regions of the country.
This document outlines the conceptual and methodological framework of the 2024 HFS, along with a description of the operational process that supports it. The structure of the document is as follows:

- Section III describes the questionnaire used in this edition and the modifications compared to the previous round.
- Section IV details the sampling design implemented.
- Section V presents the fieldwork results.
- Sections VI, VII, and VIII address technical aspects related to the calculation of expansion factors, variance estimation, and the treatment of missing data, respectively.

[1] For example, in Spain (see Bover, 2004, 2008) and in the United States (see Kennickell and Woodburn, 1997).

III. EFH 2024 Questionnaire

This section presents the **2024 Household Financial Survey (HFS) Questionnaire**, beginning with the identification of the unit of analysis—the household—and the respondent within each household. It then describes in detail the modules that comprise the questionnaire and summarizes the main changes introduced in this edition compared to the previous round.

III.1. Definition of Household and Respondent

In the HFS, a household is defined as[2]:

i) a group of individuals living in the same dwelling and sharing a food budget, or
ii) a person living alone and self-providing food.

Individuals who are temporarily absent from the dwelling are still considered part of the household, provided that their absence does not exceed six months. This includes, for example, people on vacation, engaged in temporary employment, studying away from home, among other similar situations.

After the initial contact with the household, the interviewer must identify the most suitable respondent to represent the household in the survey. The primary target is the household head. If the head is unavailable, the interviewer should seek another member with the most comprehensive knowledge of the household's finances, including income, expenditures, and debts. If this person is also unavailable and the household holds debts, the respondent should be the individual with the highest outstanding debt. If the household has no debts, the person with the largest financial and/or real estate assets should be selected. If none of the above are available, the respondent should be the individual contributing the highest income to the household. In all cases, the selected respondent must be at least 18 years old.

For the panel sample, it must be verified whether the respondent from the previous wave still resides in the household. If available, the same individual should be retained as the respondent. Otherwise, a new qualified respondent must be identified, following the criteria outlined above.

[2] The definition of household used in the HFCS follows international standards. See Organisation for Economic Co-operation and Development (OECD), *Glossary of Statistical Terms* (https://stats.oecd.org/glossary).

III.2. Topics of the Questionnaire

The 2024 HFS questionnaire consists of **15 thematic modules**, which collect information at different levels: individual (for each household member), household-level, or exclusively from the selected respondent. The following table summarizes the name of each module, and the corresponding level of information collected:

Table 1

Description of questionnaire modules and level of information collected

Module	Level of information
A. Household composition	Individual
B. Education	Individual
C. Employment status	Individual
D. Payment instruments	Household
E1. Real assets and mortgage debt (primary residence)	Household
E2. Real assets and mortgage debt (other properties)	Household
F. Non-mortgage debt	Household
G. Financial burden and credit constraints	Household
H. Vehicles and other real assets	Household
I. Financial assets, pensions, and insurance	Household /Respondent
J. Related income	Individual
L. Other income / Unrelated income	Household
K1. Expectations, perceptions, and financial literacy	Respondent
K2. Household financial decision-making	Household
K3. Expenditures	Household

Below is a brief description of each module:

A. Household Composition and Respondent Identification

Identifies all household members, the household head, and their relationships. Information is collected on gender, age, nationality, and marital status (only for individuals aged 15 and older).

B. Education

Collects educational background for all household members aged 15 and older. The respondent provides answers representing all members. Questions include the highest level of education attained, current school attendance, and tuition payments.

C. Employment Status

Explores the labor status of each household member aged 15 and older. It includes questions on employment during the previous week, unemployment or inactivity, secondary paid activities, retirement, duration of unemployment, characteristics of current employment, and pension system affiliation.

D. Payment Instruments

Investigates the ownership and use of payment instruments at the household level: checking accounts, debit cards, bank and non-bank credit cards, cash, checks, online payments, among others.

E. Real Assets and Mortgage Debt

Divided into two submodules:
- **E1**: **Primary residence** – Type, ownership, purchase date, purchase price, estimated value, financing structure, subsidies, loans, and characteristics of current mortgage.
- **E2**: **Other properties** – Up to three properties with the highest value. Includes acquisition year, estimated value, financing, and associated debt.

F. Non-Mortgage Debt

Includes credit cards, credit lines, consumer loans, auto loans, student loans, secured loans, among others. The three largest debts are identified along with their characteristics.

G. Perceived Financial Burden and Credit Constraints

Assesses the household’s perception of its debt level, ability to handle financial shocks, instances of missed payments, and access to credit.

H. Vehicles and Other Real Assets

Quantifies assets such as vehicles, machinery, artwork, livestock, among others.

I. Financial Assets, Pensions, and Insurance

Characterizes ownership and amounts in financial instruments, pension plans, and insurance policies, including informal savings, saving habits, and voluntary insurance. Also includes investment flows over the past year to capture recent financial activity.

J. Related Income

For household members with paid employment or aged 15 and older. Monthly income is reported for main and secondary occupations.

L. Other Income / Unrelated Income

Captures income not related to paid employment: pensions, subsidies, capital gains, severance payments, rental income, among others.

New Modules Introduced in the 2024 HFS:

K1. Expectations and Perceptions

This module is addressed exclusively to the respondent and aims to gather information on their expectations and perceptions regarding scenarios that may affect their personal finances. It also includes internationally standardized questions to assess financial literacy.

- **Labor expectations:** Employed respondents are asked about the likelihood of losing their job before a specific future date; job seekers are asked about the likelihood of finding employment within the same timeframe. Self-employed individuals are asked about the probability of being unable to continue their main activity due to external factors.
- **Retirement expectations:** Respondents are asked at what age they expect to retire permanently.
- **Risk aversion and time preference:** Includes experimental questions on willingness to accept delayed lottery winnings and risk-taking in uncertain scenarios (e.g., coin toss games with varying payouts). These questions measure risk tolerance and intertemporal preferences.
- **Emergency preparedness:** Respondents are asked whether the household could gather the equivalent of six months of income in case of emergency, and what sources they would rely on (e.g., savings, banks, family, asset sales).
- **Financial literacy (Big3):** The module concludes with three standard questions on compound interest, inflation, and risk diversification, widely used in international literature to assess basic financial knowledge.

K2. Household Financial Decision-Making

This module applies only to multi-person households (i.e., those with two or more members). It aims to collect information on the distribution of roles and financial decision-making dynamics within the household, from a structural perspective not tied to specific events.

- **Questions K12 and K13:** Explore how household income is managed—whether all income is pooled, partially shared, or individually managed. This helps assess the proportion of income considered "household income" and analyze debt dynamics and financial vulnerabilities.
- **Question K14:** Examines how the household handles payment issues related to individual debts, distinguishing between individual solutions, partial support, or shared responsibility.
- **Questions K15 and K16:** Focus on decision-making for everyday purchases and high-value items (e.g., furniture, vehicles, appliances), identifying who makes the decisions (respondent, partner, or jointly).
- **Questions K17 to K19:** Directed at couples leading the household, these questions address decision-making on debt and use of shared savings, allowing analysis of co-responsibility and potential gender biases in financial management.

K3. Expenditure

The main objective of this module is to collect detailed information on household expenditures—a dimension previously absent from the HFS. Combining consumption data with financial information will improve understanding of:

i. household responses to economic shocks or policy changes affecting their budgets, and
ii. household financial vulnerability.

Respondents are asked to report household spending across various categories and timeframes to minimize unintentional omissions. The question structure is inspired by international instruments such as EPS (Spain), HILDA (Australia), and HFCS (Europe). The categories defined for the 3- and 6-month horizons are subsets of those used in Chile's Household Budget Survey (EPF).

- **Average monthly spending:** Includes supermarket purchases, local markets, neighborhood stores, and basic services such as water, electricity, gas, firewood, internet, paid TV, phone services, and building/condominium fees.
- **Transportation expenses:** Covers average monthly spending on public transport, fuel, and tolls/electronic tags.
- **Health expenses:** Total spending over the past 3 months on medical consultations (including dental, psychological, physical therapy, etc.), outpatient procedures/tests, and medications.
- **Durable goods and other categories:** Total spending over the past 3 months on small appliances, home maintenance materials/services, clothing and footwear, and mobile devices (K25). For the past 6 months, spending on large appliances and furniture is also recorded.
- **Income-expenditure balance:** Respondents are asked whether household spending over the past 6 months was greater than, equal to, or less than income. If spending exceeded income, they are asked how the gap was covered (e.g., borrowing, asset sales, savings, family support, increased income, or reduced spending).

III.3. Questionnaire Modifications

As part of the continuous improvement process of the HFS, each survey wave includes a comprehensive review of both the questionnaire, and the procedures associated with its implementation. Typically, this process begins with the questionnaire used in the previous wave, which is then modified as needed. However, since the 2021 edition required adaptation and reduction for telephone-based implementation, the development of the 2024 questionnaire used both the 2017 and 2021 versions as reference points.

The review and analysis of the questionnaires focused on two main dimensions: first, the examination of non-response rates and the quality of responses obtained in the 2017 and 2021 waves; second, the identification of questions that could be removed to allow for the inclusion of the new modules K1, K2, and K3, while maintaining an average interview duration comparable to previous rounds.

Below are the main changes by module:

A. Household Composition and Respondent Identification

This module remains largely unchanged from the 2017 version. The nationality options were updated in line with changes introduced in the 2021 wave, and a new question was added to capture the relationship between each household member and the respondent. This enhancement improves the identification of family ties and facilitates analysis of household structure.

B. Education

This module remains unchanged from the 2017 version. The same questions and structure are retained.

C. Employment Status

This module follows the structure used in 2017, with minor adjustments. A new question was added to determine whether the household member has ever worked, targeting individuals who are currently neither employed nor retired. This helps filter job search questions for those who have never worked. As in the 2021 version, the question on weekly working hours is included. Additionally, the scope of certain questions was adjusted to ensure consistency with the nature of the information collected, extending criteria already applied in 2021:

1. The question on employment contract type is not asked to employers or self-employed workers.
2. The question on issuing invoices is asked only to employers, self-employed individuals, and dependent workers without a formal contract.

D. Payment Instruments

This module retains the questions that were removed in 2021 compared to the 2017 version, specifically those related to the frequency and reasons for using each payment method.

The three payment instruments added in 2021 are also retained: online payments (e.g., PayPal, Webpay, MercadoPago), physical or virtual prepaid cards, and cryptocurrency-linked cards.

Questions related to digital financial services were removed to optimize the module's length and relevance.

E. Real Assets and Mortgage Debt

For both the primary residence (E1) and other properties (E2), the 2024 questionnaire reinstates the structure and content of the 2017 version, recovering questions and formats that were removed or adjusted in 2021 due to the survey mode.

Compared to the 2017 version, the following changes were introduced:

- A new question was added for renters to estimate the market value of the primary residence, whereas previously this was asked only of owners.
- Questions regarding the name of the bank or institution providing the mortgage loan were removed for both primary and other properties (a change already implemented in 2021).
- The question on whether the mortgage was denominated in pesos or UF was removed for both property types (also implemented in 2021).
- A new question was added on the type of interest rate (fixed, variable, or mixed) for both primary and other properties.
-

F. Non-Mortgage Debt

The 2024 version of this module reinstates the structure and content of the 2017 version, recovering questions and formats that were removed or adjusted in 2021. Questions regarding the name of the bank or institution associated with credit cards, credit lines, and loans were removed for all types of debt, change already implemented in 2021.

G. Perceived Financial Burden and Credit Constraints

This module remains largely unchanged from the 2017 version. Two new questions were added:

1. The reasons why the household faced unexpected expenses in the past 12 months.
2. The amount spent to cover such expenses.

H. Vehicles and Other Real Assets

No changes were made to this module compared to the 2017 version.

I. Financial Assets, Pensions, and Insurance

The 2024 version reinstates the structure and content of the 2017 edition, recovering questions and formats that were removed or adjusted in 2021 due to the survey mode.

Questions added:

- A question on the amount invested in each type of financial asset over the past 12 months, to better characterize household investment behavior.
- A question on the balance held in checking or debit accounts, to capture more detailed information on household liquidity.

Questions removed:

- Questions about the financial institution managing the asset and about other financial assets.
- Questions on saving frequency, pension fund ownership, and insurance ownership. These removals were already implemented in 2021 and are consolidated in the 2024 version.

J. Related Income

No changes were made to this module compared to the 2017 version.

L. Other Income / Unrelated Income

No changes were made to this module compared to the 2017 version.

IV. Sample Design

The use of probabilistic sampling in survey-based statistical research ensures strong statistical properties in the results. The fact that each unit in the population has a known and non-zero probability of being selected allows for representative samples and valid inferences about the target population.

IV.1. Sample Design Characteristics

The target population of the 2024 HFS consists of urban households residing in private dwellings across the national territory. The survey has national urban coverage, with representativeness at the level of geographic macrozones[3] and wealth strata[4], defined according to the fiscal valuation of the dwelling.

As in the previous wave, the 2024 HFS was designed as a two-period rotating panel. The sample is composed of two groups: households previously interviewed in the 2021 wave (panel sample) and households interviewed for the first time in 2024 (refreshment sample).

The sample design has two main objectives:

1. To ensure a random and representative sample of the urban population, incorporating oversampling of high-wealth households to improve the precision of key variable estimates.
2. To maintain the panel component by including households that participated in the 2021 HFS, thereby preserving representativeness of the target population over time, especially for variables of interest.

The sample design of the 2024 HFS, as in previous editions, is probabilistic, stratified, and three-stage. Strata are defined by the intersection of macrozones and wealth strata.

- **First stage:** selection of municipalities.
- **Second stage:** selection of blocks within selected municipalities.
- **Third stage:** selection of dwellings within each selected block.

[3] Macrozones in the HFS refer to regional groupings used for sample design and analytical purposes. The Northern macrozone includes Regions XV, I, II, III, and IV; the Central macrozone includes Regions V, VI, VII, XVI, and VIII; the Southern macrozone includes Regions IX, XIV, X, XI, and XII; and the Metropolitan macrozone (RM) corresponds to the Santiago Metropolitan Region.

[4] This variable corresponds to a classification based on the distribution of the fiscal valuation of properties, obtained from the sampling frame (Real Estate Cadastre). The sampling frame was divided into three wealth strata according to the fiscal valuation: Stratum 1 includes percentiles 1 to 50, Stratum 2 includes percentiles 51 to 80, and Stratum 3 includes percentiles 81 to 100.

This approach ensures geographic and socioeconomic representativeness and allows for oversampling of high-wealth households. Additionally, the design is comparable to that used in the **Survey of Consumer Finances (SCF)** in the United States and the **Household Finance and Consumption Survey (HFCS)** coordinated by the European Central Bank.

The target sample size for the 2024 HFS remains the same as in the 2021 and 2017 waves, with approximately 4,500 households, of which 2,800 correspond to the refreshment sample and 1,700 to the panel sample. The characteristics of each sample type are described below.

IV.2. Panel Sample

The panel sample of the 2024 HFS consists of the 2,855 households that participated as part of the refreshment sample in the 2021 HFS. During the last quarter of 2023, a contact process was carried out with these households to thank them for their participation in the 2021 HFS and to verify their willingness to participate in the 2024 wave. As a result of this process, 432 households could not be contacted, 68 households declined the interview, and 195 households updated their contact information but ultimately refused to participate in the 2024 HFS. Thus, the available panel sample for the 2024 HFS consists of 2,155 households, with the goal of achieving approximately 1,700 completed interviews (see Table 2).

Table 2

Target panel sample by wealth strata and macrozones

Macrozones	Wealth stratums			Total
	1	2	3	
North	75	53	86	**214**
Center	185	132	210	**527**
South	208	149	238	**595**
Metropolitan región (RM)	127	91	146	**364**
Total	**595**	**425**	**680**	**1,700**

Since the 2024 panel sample is derived from the 2021 refreshment sample, it inherits the sample design of the latter. For further details on the original design, please refer to the 2021 HFS methodological report.

IV.3. Refreshment Sample

As previously mentioned, following the methodology used in earlier HFS waves, the design of the 2024 refreshment sample is probabilistic, stratified, and three-stage. The strata are defined by the intersections of macrozones and wealth strata. Second first stage, municipalities are selected; in the second stage, blocks are randomly selected within those municipalities; and in the third and final stage, dwellings are randomly selected within each block.

IV.3.1. Sampling Frame

The sampling frame for the HFS was constructed using the **Real Estate Cadastre (CBR)** from the **Internal Revenue Service (SII)**, updated as of December 2023. This frame has national coverage and includes information on fiscal valuation, municipality, block, and address for approximately 8.1 million properties. Of these, 5.7 million are residential and are referred to as dwellings.

Additionally, population projections from the **National Statistics Institute (INE)**, based on the 2017 census and updated to December 2023[5], were used to identify predominantly urban municipalities—151 out of the country's 346 municipalities. Once these urban municipalities[6] were identified in the 2023 CBR, the final sampling frame for the 2024 HFS consisted of **4,993,318 dwellings** that meet the criteria defined for the target population.

To avoid interviewing households that participated in previous waves, blocks selected in both the 2017 and 2021 rounds were excluded from the 2024 refreshment sample selection process.

IV.3.2. Stratification

One of the key attributes of the CBR is the inclusion of the **fiscal valuation** of properties, which serves as a proxy for household wealth. This enables the definition of strata and the oversampling of high-wealth households.

To implement the oversampling, dwellings in the CBR were stratified by fiscal valuation. The classification was based on the distribution of fiscal valuations within each macrozone:

- **Stratum 1**: percentiles 1 to 50
- **Stratum 2**: percentiles 51 to 80
- **Stratum 3**: percentiles 81 to 100

IV.3.3. Sample Size

The target sample size for the 2024 refreshment sample was set at approximately **2,800 households**, distributed by macrozone and wealth stratum. The precision of estimates was evaluated using **bank**

[5] https://www.ine.gob.cl/estadisticas/sociales/demografia-y-vitales/proyecciones-de-poblacion.

[6] Urban municipalities are defined as those in which less than 25% of the population is classified as rural, according to the Population and Housing Census.

debt ownership as the key variable and **relative error**[7] as the criterion, both at the total level and within each cell formed by the intersection of macrozones and wealth strata.

The analysis included a **100% oversampling** of households in the highest wealth stratum and an increased share for the **Southern macrozone**, to improve precision for that domain. **Table 3** shows the distribution of the refreshment sample by macrozone and stratum.

Table 3

Distribution of the 2024 HFS Refreshment Sample by Macrozone and Wealth Stratum

(Households and approximate percentage of total sample)

Macrozones	Estratum			Total
	1	2	3	
North	132	89	150	371
	4.7%	*3.2%*	*5.3%*	*13.2%*
Center	337	251	362	950
	12.0%	*8.9%*	*12.9%*	*33.7%*
South	184	130	207	521
	6.5%	*4.6%*	*7.3%*	*18.5%*
RM	356	238	381	975
	12.6%	*8.4%*	*13.5%*	*34.6%*
Total	1,009	708	1,100	2,817
	35.8%	*25.1%*	*39.0%*	*100.0%*

*Cifras secundarias en porcentaje aproximadas.

Table 4 presents the expected relative error for the refreshment sample, disaggregated by macrozone and fiscal valuation stratum[8]. The results indicate that the relative error for the total sample is 8.0%, which, according to the classification by INE (2025), is considered “Very good.”

[7] Relative error is the maximum percentage deviation expected between the estimated parameter and the true population parameter, with a probability of 1−α1 - \alpha1−α. This statistic allows for the standardization of measurement units across various indicators, as it is expressed as a percentage relative to the estimator.
The formula used to calculate relative error in the case of a proportion is:

$$e = 1{,}96\sqrt{\frac{deff * p(1-p)}{Np^2}}$$

Where:

- **deff** is the design effect,
- **p** is the point estimate of the reference variable,
- **N** is the sample size being evaluated, and
- **1.96** corresponds to the percentile of the normal distribution associated with a 95% confidence level (1−α1 - \alpha1−α).

For further details on the calculation and use of relative error, see INE (2025).

[8] Annex 1 presents the expected relative error for strata generated from the intersection of wealth stratum and macrozone.

By macrozone, the indicator is classified as "Good" (below 20%), with the Northern macrozone showing the highest value at 18.5%. By stratum, Strata 1 and 2 exhibit relative errors classified as "Good," while Stratum 3 achieves a "Very good" rating. This latter result is associated with the oversampling applied to this stratum, which allows for better representation of phenomena within this segment of the population.

Table 4

Expected Relative Error for the 2024 Refreshment Sample by Macrozone and Wealth Stratum

Macrozone	Estratum	Bank Debt Ownership 2021	Design Effect (2021)	Sample Size (2024)	Relative Error (2024)
1. North		37.7%	1.99	371	18.5%
2. Center		31.1%	2.18	950	14.0%
3. South		34.3%	1.16	521	12.8%
4. RM		36.9%	3.52	975	15.4%
	Estratum 1	25.3%	2.72	1,009	17.5%
	Estratum 2	40.1%	2.21	708	13.4%
	Estratum 3	51.6%	1.31	1,100	6.5%
TOTAL		**34.9%**	**2.53**	**2817**	**8.0%**

IV.3.4. First Sampling Stage: Municipality Selection

Between 2007 and 2014, the sample of municipalities remained unchanged. In 2017, it was updated based on demographic and operational criteria to improve sampling efficiency. In 2021, an additional modification was introduced to ensure two key technical aspects:

- Randomness in the selection of municipalities (Primary Sampling Units – PSU)
- Direct expansion factors for the first sampling stage

For the 2024 wave, the design implemented in 2021 was retained, with a target sample of **77 municipalities**, selected from the universe of **151 urban municipalities**.

Urban municipalities were divided into two types:

- **Self-represented municipalities**: regional capitals and conurbations (62 municipalities)
- **Co-represented municipalities**: the remaining urban municipalities (89 municipalities)

The 62 self-represented municipalities were **forcibly included** in the sample (selection probability = 1). To reach the target of 77 PSUs, the remaining 15 co-represented municipalities were distributed proportionally across macrozones. **Table 5** shows the final distribution.

Table 5

Distribution of Municipalities in the 2024 HFS Refreshment Sample by Macrozone

(Number of municipalities)

Macrozone	Urban Municipalities					
	Self-represented		Co-represented		Total	
	Population	Sample	Population	Sample	Population	Sample
North	7	7	15	3	**21**	**10**
Center	15	15	53	9	**68**	**24**
South	5	5	14	2	**19**	**7**
RM	35	35	7	1	**42**	**36**
Total	**62**	**62**	**89**	**15**	**151**	**77**

Municipality selection within each macrozone was conducted independently using Probability Proportional to Size (PPS) sampling without replacement, based on the number of dwellings. The Hanurav-Vijayan algorithm was used for selection.

The final sample of municipalities covers 4.1 million dwellings, representing approximately 83% of urban dwellings nationwide.

Details of the selected municipalities for both the refreshment and panel samples are provided in Annex 2.

IV.3.5. Second Sampling Stage: Block Selection

The secondary sampling unit (SSU) is defined as a **sub-block**, i.e., a group of dwellings within the same geographic block that belong to the same wealth stratum. If a block contains dwellings from multiple strata, it is split into separate sub-blocks. For simplicity, these are referred to as **blocks** or **SSUs**.

Starting from the target sample of 2,800 dwellings, the following conditions were applied for SSU selection:

- Each block must contain **4 dwellings** to allow for variance estimation.
- At least **one block** must be selected in each macrozone-municipality-stratum intersection.

Initial allocation of blocks per macrozone and stratum was calculated using adjusted weights for oversampling in Stratum 3 and the Southern macrozone:

$$m_{0,z,e} = w_{0,z,e} * \frac{n_0}{4}$$

Where:

n_0: is the initial target sample size (2,800 dwellings)

$w_{0,z,e}$: is the initial weight for macrozone z and stratum e.

The following tables describe the initial weights and sample sizes for block selection.

Table 6

Initial Weight Allocation by Macrozone and Wealth Stratum (% of total)

Zone	Estratum 1	Estratum 2	Estratum 3	Total
North	4.6%	3.3%	5.3%	13.2%
Center	11.7%	8.3%	13.3%	33.3%
South	6.5%	4.6%	7.4%	18.4%
RM	12.3%	8.7%	14.0%	35.0%
Total	**35.0%**	**25.0%**	**40.0%**	**100.0%**

Table 7

Initial Block Sample Allocation by Macrozone and Wealth Stratum

Zona	Estratum 1	Estratum 2	Estratum 3	Total
North	32	24	37	93
Center	82	58	93	233
South	45	32	52	129
RM	86	61	98	245
Total	**245**	**175**	**280**	**700**

Block allocation within municipalities was proportional to the number of blocks in each municipality-stratum intersection, with a minimum of one block per intersection:

$$m_{0,z,c,e} = round\left(m_{0,z,e} * \frac{M_{z,c,e}}{\sum_{\forall c} M_{z,c,e}}\right)$$

Where:

$m_{0,z,c,e}$: initial block sample size for macrozone z, municipality c and stratum e.

$M_{z,c,e}$: total number of blocks in macrozone z, municipality c and stratum e.

Final block sample size was determined as:

$$m_{z,c,e} = \begin{cases} 1 & ; \quad si\ m_{0,z,c,e} = 0 \\ m_{0,z,c,e} & ; \quad si\ m_{0,z,c,e} > 0 \end{cases}$$

The final sample includes **725 blocks**. Systematic random sampling with PPS was applied, using **average fiscal valuation** as the sorting variable within each municipality. This ensured proportional selection and avoided concentration in high-value areas.

Approximately **2,700 blocks** from the 2017 and 2021 waves were excluded to prevent re-selection. Adjustments were made to expansion factors to maintain full coverage of the target population.

Table 8
Distribution of Blocks (SSUs) in the 2024 HFS Refreshment Sample by Macrozone and Wealth Stratum

Macrozone	Stratum of wealth						Total	
	1		2		3			
	Blocks	%	Blocks	%	Blocks	%	Blocks	%
North	33	*4.6%*	23	*3.2%*	38	*5.2%*	94	*13.0%*
Center	85	*11.7%*	63	*8.7%*	94	*13.0%*	242	*33.4%*
South	47	*6.5%*	33	*4.6%*	53	*7.3%*	133	*18.3%*
RM	89	*12.3%*	64	*8.8%*	103	*14.2%*	256	*35.3%*
Total	254	*35.0%*	183	*25.2%*	288	*39.7%*	725	*100.0%*

IV.3.6. Third Sampling Stage: Dwelling Selection

The third stage corresponds to the selection of **Ultimate Sampling Units (USUs)**—the dwellings. These were selected using **systematic random sampling** within each of the 725 blocks, previously ordered by fiscal valuation. Four dwellings were selected per block; if a block had four or fewer dwellings, all were selected.

Thus, the final refreshment sample for the 2024 HFS consists of **2,817 dwellings**.

V. Fieldwork

The data collection for the 2024 HFS was awarded to **IPSOS Chile** through a public bidding process.

V.1. Pilot Test

To prepare for the 2024 HFS fieldwork, a pilot test was conducted to replicate the full survey implementation process. This exercise allowed for the evaluation of questionnaire flow, consistency, and length; the review of training materials and support resources for interviewers; and the testing of both the contact protocol and the replacement procedure for the refreshment sample. The ultimate goal was to anticipate and address potential challenges during the main data collection phase.

The pilot test was carried out in May, with the participation of 24 interviewers. A total of 109 households were surveyed, of which 97 belonged to the refreshment sample and 12 to the panel sample.

Key findings from the pilot test included:

1. The number of interviewers with the necessary skills and motivation to conduct a survey of the HFS's complexity—both in sampling and thematic scope—is limited. Most interviews were conducted by only four interviewers.
2. Household willingness to participate in surveys has declined compared to previous HFS waves. In addition to traditional challenges such as:
    a. the sensitive nature of the survey topics, and
    b. the length of the questionnaire,
    other factors were observed, including:
    c. safety concerns and distrust in opening the door to strangers, and
    d. reduced sense of collaboration with institutions.

As a result of the pilot, the following actions were implemented to improve fieldwork:

1. **Enhancement of interviewer support materials** to increase credibility and encourage household participation. The additional measures introduced for the 2024 wave included:
    a. Letters sent to municipalities requesting their support in publicizing the presence of interviewers via their official websites and social media.
    b. A notice published on the Central Bank's homepage, with a link to the HFS section for more information and interviewer verification.
    c. Improved presentation letters, including a watermark to enhance authenticity.
    d. Training the Bank's call center staff to address inquiries from the public. For non-business hours, a recorded message was implemented confirming the survey's

existence and providing guidance on how to obtain more information via the Bank's website.

e. Dissemination of relevant information about the survey on the Bank's social media channels.

2. A review of the questionnaire was conducted, eliminating questions and optimizing flow to achieve a **median interview duration of 35 minutes**.

V.2. Fieldwork Results

Fieldwork for the 2024 HFS was conducted between July 22, 2024, and January 13, 2025, spanning a total of 26 weeks.

- For the panel sample, the target was at least 1,700 interviews. A total of 2,080 households were visited out of the 2,155 that had updated their contact information and expressed interest in participating. As a result, 1,801 interviews were completed.
- For the refreshment sample, the target was 2,817 interviews. The random sample included up to four replacements per original unit, managed by the Central Bank according to the established protocol. In total, 4,724 households were visited, resulting in 2,848 interviews.

A total of 141 interviewers participated in the fieldwork, with an average of 33 interviews per interviewer. However, performance was highly uneven:

- 10 interviewers conducted 52% of all interviews, averaging 240 interviews each,
- while 65 interviewers conducted 10 or fewer interviews over the 26-week period.

The overall response rate was 71.5%, with differences between samples:

- Panel sample: 87.2%
- Refreshment sample: 64.1%

The refusal rate was 11.3%, also varying by sample:

- Refreshment sample: 14.7%
- Panel sample: 3.7%

V.3. Interview Duration

The 2024 HFS questionnaire contains a total of 855 questions. However, due to the instrument's filter structure, no household answered all of them. The household that answered the most questions completed approximately 40% of the questionnaire (around 340 questions). On average, households answered 240 questions, equivalent to 28% of the full questionnaire.

- Average interview duration: 33 minutes

- Median interview duration: 29.6 minutes
- Only 10% of households had interviews exceeding 55 minutes, with the longest interview lasting 2.5 hours.

Table 9

Average and Median Interview Duration by Module – HFS 2024 (in minutes)

Module		**Number of questions**	**Average duration**	**Median duration**
A	Household Composition	7	2.0	1.3
B	Education	5	1.2	0.7
C	Employment Status	15	4.0	2.9
D	Payment Instruments	23	2.1	1.5
E1	Real Assets and Mortgage Debt (Primary Residence)	50	2.8	1.8
E2	Real Assets and Mortgage Debt (Other Properties)	137	1.3	0.3
F	Non-Mortgage Debt	342	4.9	1.7
G	Financial Burden and Credit Constraints	33	2.0	1.0
H	Vehicles and Other Real Assets	19	1.2	0.5
I	Financial Assets, Pensions, and Insurance	64	3.3	2.2
J	Related Income	10	1.3	0.8
L	Other / Unrelated Income	81	2.0	1.4
K1	Expectations, Perceptions, and Financial Literacy	26	4.3	3.2
K2	Household Financial Decision-Making	8	1.6	0.9
K3	Expenditures	27	4.6	3.6
M	Contact Information	8	3.0	1.6
Total		**855**	**33.0**	**29.6**

VII. Weights

Probabilistic sample designs, such as the one used in the 2024 HFS, allow for the calculation of sampling errors, confidence intervals, and other key statistical indicators to assess the precision of survey results. They also provide the methodological foundation for the application of **weights**, which adjust sample estimates to accurately reflect the structure of the total population, ensuring the validity and comparability of the indicators produced.

In their simplest form, weights correspond to the inverse of the inclusion probability for each sampled unit. Their purpose is to enable unbiased estimation of population parameters of interest, making them a fundamental component of data analysis—especially when selection probabilities vary due to the complexity of the sample design.

The 2024 HFS employs a complex sample design, involving two independent samples (panel and refreshment), and combining different sampling methods (stratified and multi-stage). Given this complexity, in addition to calculating design weights (i.e., direct weights based on the sample design), the following steps are required:

- Adjusting design weights for non-response and coverage.
- Integrating weights from both sample components.
- Calibrating weights to the target population, in accordance with the sample design.

The following sections describe the construction process of weights for both the refreshment sample and the panel sample.

VII.1. Weights for the Refreshment Sample – HFS 2024

The sample design for the refreshment sample in the 2024 HFS is probabilistic, stratified, and three-stage. Strata are defined by the intersection of geographic macrozones and wealth strata, while the sampling stages correspond to the selection of municipalities, blocks, and dwellings.

VII.1.1. First Sampling Stage: Municipalities

For the random selection of municipalities, the technique of Probability Proportional to Size (PPS) sampling without replacement was used, where the selection probability is determined by the number of dwellings in each municipality.

The probability of selecting municipality c_h in the macrozone m and the wealth stratum e is define as[9]:

$$P(c_h) = \begin{cases} 1 & ; \quad if \ \ c_h \ is \ self-represented \ (A) \\ n^B_{m,e} \dfrac{N_{h,m,e}}{\sum_{\forall h \in B} N_{h,m,e}} & ; \quad if \ \ c_h \ \ is \ co-represented \ (B) \end{cases}$$

Where:

$N_{h,m,e}$: number of dwellings in municipality h, macrozone m, and stratum e.
$n^B_{m,e}$: number of co-represented municipalities to be selected in macrozone m and stratum e.

VII.1.2. Second Sampling Stage: Blocks

Blocks were selected using systematic random sampling with probability proportional to size (PPS_SYS), where the sorting variable was the average fiscal valuation of each block within each macrozone–municipality–wealth stratum cell.

The probability of selecting block m_i in the municipality h and the wealth stratum e is defined as[10]:

$$P(m_i) = n_{h,e} \cdot \frac{\bar{x}_{i,h,e}}{\sum_{\forall i}^{N_{h,e}} \bar{x}_{i,h,e}}$$

Where:

$n_{h,e}$: number of blocks to be selected in municipality h and stratum e.
$\bar{x}_{i,h,e}$: average fiscal valuation of the dwellings in block i.
$N_{h,e}$: total number of blocks in municipality h and stratum e.

The selection procedure was implemented using the Hanurav -Vijayan algorithm[11].

[9] Although municipalities are selected without replacement, for descriptive purposes the formulas are presented as if selection were with replacement, due to the complexity involved in representing the exact selection probabilities under the without-replacement scheme.
[10] Although blocks are selected without replacement, for descriptive purposes the formulas are presented as if selection were with replacement, due to the complexity involved in representing the exact selection probabilities under the without-replacement scheme.
[11] pp. 221–234. Principles of Social Research Methodology, chapter titled Sampling Techniques for Quantitative Research.

VII.1.3. Third Sampling Stage: Dwellings

Dwellings were selected using simple random sampling without replacement, with four dwellings selected per block. If a block contained four or fewer dwellings, all were selected.

The probability of selecting dwelling v_j in the block i is given by:

$$P(v_j) = \frac{n_i}{N_i}$$

Where:

N_i: total number of dwellings in block i.
n_i: number of dwellings selected in block i (If $N_i \leq 4$, then $n_i = N_i$).

VII.1.4. Overall Selection Probability and Design Weight

The overall probability of selecting a dwelling in the refreshment sample is the product of the probabilities from each sampling stage:

$$\pi_j = P(c_h)P(m_i)P(v_j)$$

The **design weight** (also referred to as the direct expansion factor) is defined as:

$$d_j = \frac{1}{\pi_j}$$

VII.1.5. Adjustment for Under- or Over-Representation

As in previous waves, the sample design included the use of replacement dwellings, which allowed interviews to be conducted in all selected municipalities and blocks. However, the target number of interviews was not met in 24 blocks, while in 40 blocks more households were interviewed than originally planned. The effective sample consisted of 2,848 households—31 more than the planned target—making it necessary to apply an adjustment for under- or over-response at the block level.

This adjustment is expressed as:

$$d_j^{NR} = d_j \frac{1}{\emptyset_i}$$

Where:

$\emptyset_i = n_i'/n_i$ is the ratio of the actual to the planned number of dwellings in block.

This adjustment ensures that the sum of the adjusted weights d_j^{NR} reproduces the total number of dwellings in the sampling frame.

Table 10
Sum of Adjusted Weights for the Refreshment Sample – HFS 2024
(By Macrozone and Wealth Stratum)

Zone	Stratum 1	Stratum 2	Stratum 3	Total
North	319,319	198,791	131,534	649,644
Center	803,623	486,676	348,776	1,639,075
South	208,037	132,855	80,290	421,182
RM	1,165,688	679,680	438,049	2,283,417
Total	2,496,667	1,498,002	998,649	4,993,318

VII.2. Panel Sample Weights

As previously mentioned, the Household Financial Survey (HFS) includes two independent sample designs. The second corresponds to the panel sample, composed of a subset of households that participated in the refreshment sample of the 2021 HFS. Consequently, the initial weight for this sample corresponds to the non-response adjusted weight from that round:

$$d_j^{NR,21} = d_j^{21} * \frac{1}{\phi_j} \qquad where \qquad d_j^{21} = \frac{1}{\pi_j^{21}}$$

As detailed in Section IV.2, the 2021 refreshment sample was revisited in 2023 through a recontact process aimed at updating the contact information necessary for survey implementation. As a result, updated information was obtained for 2,350 households (82% of the 2021 refreshment sample), of which 2,155 households (76%) expressed willingness to participate in the new wave. Ultimately, during the 2024 HFS fieldwork, 1,801 households (63% of the 2021 refreshment sample) were successfully interviewed, forming an effective panel sample.

The phenomenon of attrition between consecutive waves requires a non-response adjustment, since households that do not participate may differ systematically from those that do (e.g., in geographic location, education level, age of the main respondent, among other characteristics). This adjustment aims to minimize bias attributable to non-random causes of non-response.

As in the previous wave, the non-response adjustment is performed using a binary logit model, which estimates conditional response probabilities based on a set of variables that characterize the 2021 refreshment sample. The general specification of the model is:

$$\Pr(R = 1 | x_j \beta) = \Lambda(x_j \beta) = \phi_j$$

From this model, predicted response probabilities $\widehat{\phi}_i$ are obtained and grouped into ventiles ($s = 1, \dots ,20$). The non-response adjusted weight is expressed as:

$$d_j^{NR\prime,21} = d_j^{NR,21} * 1/\bar{\phi}_s$$

The weight $d_j^{NR',21}$ serves as the base weight for the panel sample. However, since it expands to the sampling frame of the previous wave, a final adjustment is required to ensure proper representation of the macrozones in the 2024 HFS sampling frame (CBR2023):

$$d_j^{NR',21,24} = d_j^{NR',21} * \frac{N_{m,e}^{24}}{\sum_{\forall j \in m,e} d_j^{NR',21}}$$

Where:

$N_{m,e}^{24}$: number of dwellings in macrozone m and stratum e in the 2023 Real Estate Cadastre (CBR2023).
$d_j^{NR',21,24}$: final panel weight, used in the construction of the cross-sectional weight for the 2024 HFS

The following table presents the sum of the sample weights of the panel sample, adjusted for coverage and non-response.

Table 11

Sum of Adjusted Weights for the Panel Sample – HFS 2024
(By Macrozone)

Zone	Total
North	649,644
Center	1,639,075
South	421,182
RM	2,283,417
Total	4,993,318

The following table presents a comparison of selected descriptive statistics for the refreshment sample weights from the 2021 HFS and the adjusted panel sample weights for the 2024 HFS. The results confirm that the sum of the panel sample weights, adjusted for non-response and aligned with the 2024 sampling frame, reproduces the total number of urban dwellings in the CBR2023.

Table 12

Impact of Adjustments on Panel Sample Weights – HFS 2024

Statistic	Refreshment 2021	Panel 2024 (NR Adjusted)	Panel 2024 (NR + Frame Adjusted)
Sum	4,486,025	4,473,941	4,993,318
N	2,855	1,801	1,801
Mean	1,571	2,484	2,773
sd	1,808	3,293	3,675
p5	216	276	316
p25	560	750	871
p50	1,255	1,716	1,929
p75	2,167	2,978	3,292
p95	3,503	7,072	7,905
max	21,921	49,260	54,477

VII.3. Cross-Sectional Weights – HFS 2024

As in the two previous waves of the HFS, the adjusted weights from each sample type are combined into a single cross-sectional weight, defined as follows:

$$w_j^{24} = \begin{cases} \alpha^R * d_j^{NR} & if \quad j \in Refreshment \\ (1-\alpha^R) * d_j^{NR',21,24} & if \quad j \in Panel \end{cases}$$

Where:

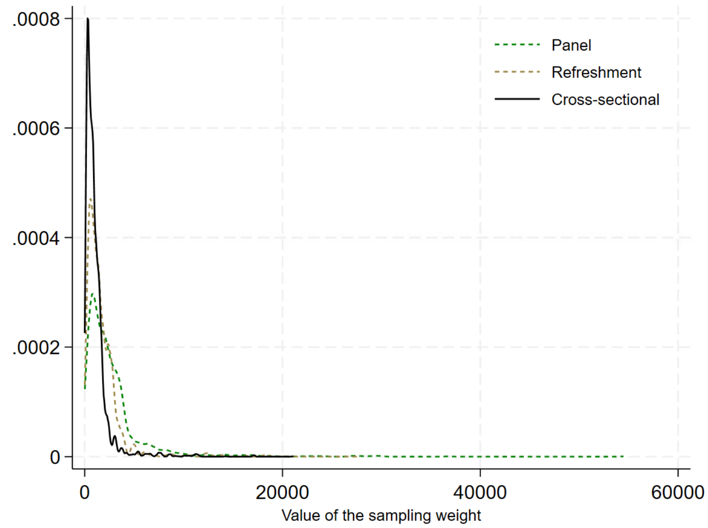
α^R: proportion of the refreshment sample relative to the total HFS 2024 sample, calculated as $\alpha^R = n^R/_n$, with $n = n^R + n^P$.

This methodology assigns a value to each sample type based on its proportion within the overall sample size. This approach has been used in the Spanish Household Financial Survey (EFF) (Bover et al., 2014) and in the U.S. Survey of Consumer Finances (SCF) (Kennickell & Woodburn, 1998).

The following figure shows the density distribution of the panel and refreshment sample weights, along with the resulting cross-sectional weights. It illustrates how combining both sample types reduces the dispersion of individual weights, ensuring that each interviewed dwelling represents fewer dwellings than it originally did under its respective sample design.

Figure 1

Distribution of Combined (Cross-Sectional) Weights – HFS 2024



VII.4. Calibration

The calibration process consists of adjusting the expansion weights—previously corrected for non-response and combined into a single cross-sectional weight—so that the resulting estimates better reflect known characteristics of the target population.

This adjustment is particularly recommended when aiming to minimize the effects of non-response and undercoverage. Once applied, it ensures that the expansion weights match population totals for key demographic variables such as age groups, gender, and geographic zones.

The population totals used to calibrate the 2024 HFS sample correspond to urban population projections for the year 2024, published by the National Statistics Institute (INE). As in the 2021 wave, the calibration of both sample types is performed jointly using the generalized regression method. The cross-sectional weights serve as the main input for this process.

Using the `calibrate` command in Stata, the cross-sectional weights (at the household level) are calibrated for each household member P_i in the household j, based on urban population projections by age, gender, and macrozone, according to the following expression:

$$w_{jP_i}^{24} = w_j^{24} * \delta_{jP_i}$$

Where:

δ_{jP_i} : calibration adjustment for individual i in household j.

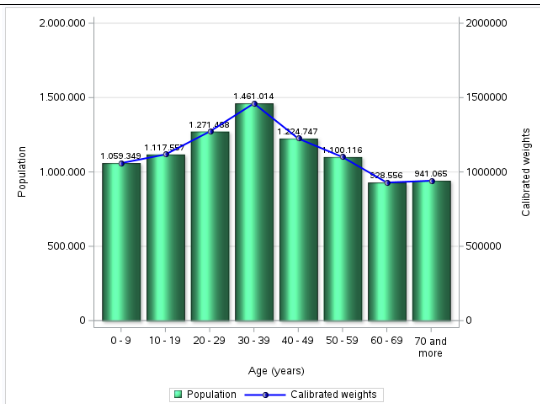
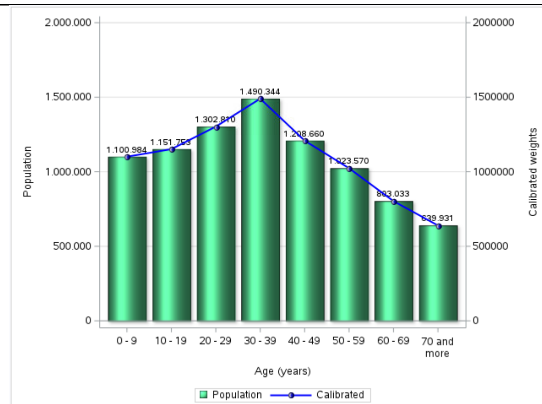
Finally, the average of the calibrated weights $w_{jP_i}^{24}$ within each household is calculated, resulting in the **final calibrated cross-sectional weight** for the 2024 HFS:

$$w_{jP_i}^{24}$$

The following charts illustrate the adjustment of calibrated expansion weights. They show that this process ensures **exact reproduction** of the projected population totals for 2024, disaggregated by gender:

Figure 2

2024 Population Projections vs. Calibrated Expansion Weights – HFS 2024

Women	Men
	

VIII. Partial Non-Response and Imputation

The sensitive and detailed nature of the information collected in the Household Financial Survey (HFS)—such as income, assets, debts, and expenditures—results in higher rates of partial non-response compared to surveys on less sensitive topics. In this context, missing values pose a significant challenge for survey design and analysis. If not properly addressed, they can introduce bias into the results and reduce the efficiency of estimates.

In surveys such as the Survey of Consumer Finances (SCF) in the United States and the Spanish Household Financial Survey (EFF), missing data are handled using Multiple Imputation by Chained Equations (MICE). This method preserves the uncertainty associated with missing values to ensure statistically valid inferences, generating multiple plausible values rather than a single point estimate. Various methods exist for inputting missing data, including Hot Deck, K-Nearest Neighbors (KNN), hierarchical Bayesian models, and Random Forest. MICE was selected for its model-based formulation using linear regressions, which facilitates implementation, interpretability, and traceability in statistical contexts. The procedure is structured as a sequence of chained regressions, designed to capture multivariate relationships between variables. This architecture is computationally demanding, as it does not allow for parallelization or intervention during execution. Validation is performed through pre-review of regression models and post-evaluation of the coherence and stability of the imputations.

At the Central Bank of Chile, this procedure is systematically applied in each wave, generating 30 imputed versions per variable, consolidated into a single dataset for analysis. Each version is obtained after 30 chained iterations, required to reach convergence, in a process that takes approximately 30 hours of computation.

Rigorous treatment of missing data has become standard practice in international financial surveys. The Chilean HFS aligns with these best practices, adopting imputation procedures that enable the generation of representative, accurate, and internationally comparable information.

VIII.1. MICE Imputation Procedure in the 2024 HFS

The implementation of the MICE method in the 2024 HFS follows these steps:

1. Identification of missing values: Variables with missing data are detected and classified by type:
    - *Individual-level variables:* e.g., income from main occupation, asked for each employed household member.
    - *Household-level variables:* e.g., mortgage debt amount or property value.
    - *Repeated variables:* asked multiple times per household, such as the three largest consumer debts.

2. Specification of imputation models: A suitable conditional regression model is defined for each variable with missing data.
3. Initialization: An initial imputation is performed to complete the dataset and allow the algorithm to begin.
4. Chained iteration: Missing values are imputed sequentially until stability is reached. In the HFS, three thematic blocks of variables are used (defined in step 1), requiring a double chained iteration process: first within each model, then across blocks.
5. Generation of multiple imputed datasets: The full process is repeated to obtain multiple complete versions of the dataset.
6. Post-imputation analysis: Analyses are conducted on each imputed dataset using Rubin's rules (see Annex 3), which allow for valid estimation of parameters and standard errors.

VIII.2. Analysis of Missing Values

Since its first implementation in 2007, HFS only impute continuous variables, such as income amounts, assets, debts, expenditures, and hours worked. In 2024, 249 imputable variables were identified, of which 222 were imputed (see Table 12).

Table 12

Imputable and Imputed Variables, and Imputation Method Used

Category	Total variables	Imputed	Hot Deck	MICE	Max % Missing	Min % Missing
Individual Income	7	7	2	5	29.7%	2.8%
Household Income	30	29	13	16	33.0%	1.4%
Household Assets	30	30	10	20	63.7%	4.8%
Mortgage Debt	53	41	24	17	71.4%	0.4%
Household Expenditure	21	21	0	21	38.3%	4.0%
Consumer Debt[12]	108	94	4	90	44.7%	6.1%
Total	**249**	**222**	**53**	**169**	**71.4%**	**0.4%**

Variables with fewer than 55 observations are imputed using the Hot Deck method, while the remaining variables are imputed using MICE. The 169 variables imputed via MICE are grouped into three thematic blocks (see Annex 4).

VIII.3. Variable Transformation and Initial Values

To impute a block using variables from another block, initial values are assigned to the missing data in external variables using the Hot Deck method. Additionally, statistical transformations are applied to improve the behavior of variables, such as Box-Cox or logarithmic transformations (see Annex 5).

[12] Aggregates of the three largest debts. For example, the variables cf_td1_1, cf_td1_2, and cf_td1_3, which correspond to the financial burden of credit card debts for the three largest debts, are grouped into a single variable (cf_td1_) for the imputation process.

Pre-imputation steps for MICE:

- **Initial imputation using Hot Deck**: All 222 variables are imputed to generate initial values.
- **Validation using response brackets**: Imputed values are adjusted according to the ranges declared by respondents.
- **Box-Cox transformation**: The bcskew0 command in Stata is used to reduce variable skewness.

VIII.4. Evaluation and Consistency of Imputations

The imputation process was evaluated and refined iteratively through the following steps:

1. **The imputation process was evaluated and refined iteratively through the following steps:**
    - *Global metrics:* R^2, adjusted R^2, F-statistic, Root MSE, and sample size.
    - *Internal structure:* analysis of regressor significance, ranking by relevance, and conceptual coherence.
2. **Consistency evaluation by economic blocks:**
    - Comparison between models of related variables.
    - Methodological imitation from successful models.
3. **Evaluation of imputed data:**
    - Comparison of statistical distributions between original and imputed data.
    - Validation using histograms, boxplots, and percentile analysis.
    - Application of Rubin's (1987) framework to consolidate statistics.

IX. Variance Estimation

As in other household financial surveys, such as the Household Finance and Consumption Survey (HFCS) coordinated by the European Central Bank, the HFS makes available to the public replicate weights (Bootstrap weights). This approach is highly effective for estimating the variance of survey estimators due to its ease of implementation and its ability to produce consistent error estimates for both continuous and categorical variables (Shao, 2003; Girard, 2009).

IX.1. Variance Estimation Methods in Complex Surveys

Variance estimation methods in complex surveys are generally divided into two main groups:

- **Parametric methods,** such as **Taylor Linearization (TLS)**
- **Replication-based (non-parametric) methods**, including:
    - *Balanced Repeated Replication (BRR)*
    - *Jackknife Repeated Replication (JRR)*
    - *Bootstrap*

Regardless of the method used, two simplifying assumptions are commonly applied:

1. The **Primary Sampling Unit (PSU)** is treated as the ultimate unit of information, allowing any complex design to be reduced to a **two-stage stratified sampling design** (Ultimate Cluster Sampling).
2. PSUs are assumed to be selected **with replacement** from first-stage strata.

These assumptions tend to produce a **slight overestimation of variances** (Heeringa et al., 2017).

IX.1.1. Taylor Linearization Method

This method involves linearizing the estimator using a Taylor series expansion and then computing its variance using exact methods. To implement this in statistical software such as Stata, it is necessary to have access to cluster, stratum, and weight variables. This requirement can be a limitation, as some surveys do not publish these variables to protect respondent confidentiality.

IX.1.2. Replication Methods

The BRR, JRR, and Bootstrap methods use replicates of the original sample to estimate the variance of survey estimators. The general procedure includes the following steps:

1. Define $r = 1, \ldots, R$ replicates of the sample.

2. Compute replicate weights by rescaling the original weights and applying **non-response and calibration adjustments**.
3. Estimate the parameter of interest using both the original and replicate weights.
4. Calculate the variance of the estimator.

The difference between methods lies in how the replicates are generated:

- **BRR**: **BRR**: Designed for designs with **two PSUs per stratum**. Each replicate retains one of the two PSUs per stratum, generating up to 2^H replicates. The **Hadamard matrix** is used to select PSUs, assigning **double the original weight** to selected PSUs and **zero** to those excluded.
- **JRR**: Each replicate **removes one PSU** from the sample. If there are 1,500 PSUs, 1,500 replicates are generated. Weights for remaining PSUs are rescaled by $n_h/(n_h - 1)$.
- **Bootstrap**: Selects $n_h - 1$ PSUs per stratum via **random sampling with replacement**. The weight for each selected PSU is calculated as:

$$\omega_{hij}^* = \frac{n_h}{n_h - 1} \cdot m_{hij}^* \cdot \omega_{hij}$$

where m_{hij}^* is the number of times PSU (h, i, j) is selected in the replicate.

IX.2. Variance Estimation in the 2024 HFS

For the 2024 HFS, the **Bootstrap method** was chosen, as it preserves household confidentiality. The public datasets do not include information on **strata or clusters**, which prevents direct application of methods requiring resampling of the original sample.

Bootstrap weights are provided as a **public resource** without compromising data privacy, making them an optimal choice for variance estimation by external users.

The procedure applied in the 2024 HFS is as follows:

1. For each stratum h, a random sample with replacement is drawn to select a sample of size $n_h - 1$.
2. Weights for each replicate are generated using the same method as for the original weights. These are denoted w_{hi}^*.
3. The above steps are repeated 1,000 times, producing $\{w_{hi}^{*b}\}_{b=1}^{1000}$ for each unit (h, i).

Given the three-stage sample design of the 2024 HFS, this method assumes that variability arises primarily from the first two stages (selection of municipalities and blocks), which is reasonable, as household-level variability tends to be lower (HFCS, 2020).

As in the 2017 and 2021 waves, methodological improvement was incorporated: instead of directly applying the Rao-Wu rescaling formula (Rao et al., 1992; Girard, 2009; HFCS, 2016), the entire process of generating original weights was replicated in each bootstrap replicate, including non-response and calibration adjustments.

X. References

Albacete, N. (2014). “Multiple Imputation in the Austrian Household Survey on Housing Wealth”. *Austrian Journal of Statistics*, 43(1), 5–28.

Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: what is it and how does it work? *International Journal of Methods in Psychiatric Research*, 20(1), 40–49.

Barceló, C. (2006). “Imputation of the 2002 Wave of the Spanish Survey of Household Finances (EFF)”. *Documentos Ocasionales N° 0603*, Banco de España.

Ben-Gal, I. (2005). Outlier detection. En Maimon, O. & Rockach, L. (Eds.), *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*. Kluwer Academic Publishers.

Bodner, T. (2008). What improves with increased missing data imputations? *Structural Equation Modeling*, 15(4), 651–675.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), 211–252.

Bover, O. (2004). “The Spanish Survey of Household Finances (EFF): description and methods of the 2002 wave”. *Documentos Ocasionales N° 0409*, Banco de España.

Bover, O. (2008). “The Spanish Survey of Household Finances (EFF): description and methods of the 2005 wave”. *Documentos Ocasionales N° 0803*, Banco de España.

Centro UC de Encuestas y Estudios Longitudinales (2016). “Diseño Muestral y Factores de Expansión de la VI Ronda EPS”. Santiago de Chile.

Deville, J., & Sarndal, C. (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, 87(418), 376–382.

Heeringa, S. G., West, B. T., & Berglund, P. A. (2017). *Applied Survey Data Analysis* (2nd ed.). Chapman and Hall, London.

HFCN (2016). “The Eurosystem Household Finance and Consumption Survey: methodological report for the second wave”. *Statistics Paper Series N°17*.

Instituto Nacional de Estadísticas – Chile (2025) “Fundamentos del Estándar para la evaluación de la calidad de las estimaciones en encuestas económicas”. Documento de trabajo, Santiago de Chile.

Kennickell, A., & Woodburn, R. (1997). “Consistent Weight Design for the 1989, 1992 and 1995 SCFs, and the Distribution of Wealth”. *Working Paper*, Board of Governors of the Federal Reserve Board.

Little, R., & Rubin, D. (2002). *Statistical Analysis with Missing Data* (2nd ed.). John Wiley & Sons, New York.

Mander, A., & Clayton, D. (1999). HOTDECK: Stata module to impute missing values using the hotdeck method. *Statistical Software Components S366901*, Boston College Department of Economics.

McCarthy, P. J. (1969). Pseudoreplication: Half Samples. *Review of the International Statistical Institute*, 37, 239–264.

Observatorio Social (2016). "Metodología de Diseño Muestral Encuesta de Caracterización Socioeconómica Nacional, Casen 2015". *Serie Documentos Metodológicos Casen N°3*.

Portela, J., & Villeta, M. (2007). *Técnicas básicas de Muestreo con SAS*. 1ª Edición. ISBN 978846866133.

Rao, J. N. K., Wu, C. F. J., & Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology*, 18(2), 209–217.

Royston, P. (2004). Multiple imputation of missing values. *The Stata Journal*, 4(3), 227–241.

Rubin, D. B. (1976). Inference and Missing Data. *Biometrika*, 63(3), 581–592.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, New York.

Rubin, D. B. (2004). Multiple Imputation After 18+ Years. *Journal of the American Statistical Association*, 91(434), 473–489.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. Chapman & Hall/CRC, New York.

Singh, A. C., & Mohl, C. A. (1996). Understanding calibration estimators in survey sampling. *Survey Methodology*, 22, 107–115.

StataCorp. (2019). *Stata 19 Multiple-Imputation Reference Manual*. College Station, TX: Stata Press.

Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112–118.

Van Buuren, S. (2012). *Flexible Imputation of Missing Data*. CRC Press.

Van Buuren, S., Boshuizen, H. C., & Knook, D. L. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine*, 18(6), 681–694.

Vijayan, K. (1968). An Exact Sampling Scheme: Generalization of a Method of Hanurav. *Journal of the Royal Statistical Society*, 30(6), 499–513.

Valliant, R., Dever, J., & Kreuter, F. (2013). *Practical Tools for Designing and Weighting Survey Samples*. Springer, New York.
von Hippel, P. T. (2020). How many imputations do you need? A two-stage calculation using a quadratic rule. *Sociological Methods & Research*.

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30(4), 377–399.

Wolter, K. M. (1985). *Introduction to Variance Estimation*. Springer, New York.

Annex 1: Relative Error of the 2024 HFS Refreshment Sample

Macrozone	Estratrm	Proportion in 2021	Design Effect (Deff)	Sample Size (2024)	Relative Error (2024)
North	1	26.0%	1.94	132	40.1%
	2	44.9%	1.80	89	30.9%
	3	55.4%	0.68	150	11.9%
Center	1	22.4%	2.18	337	29.3%
	2	34.8%	2.14	251	24.8%
	3	48.1%	1.63	362	13.7%
South	1	23.8%	1.22	184	28.6%
	2	42.2%	0.83	130	18.3%
	3	49.6%	0.59	207	10.5%
RM	1	27.6%	3.89	356	33.2%
	2	41,8%	2.68	238	24.5%
	3	53.4%	1.41	381	11.2%
TOTAL		**34.9%**	**2.53**	**2817**	**8.0%**

Annex 2: List of Municipalities – HFS 2024

Macrozone	Municipalities	Refreshment 2024	Panel (after recontact)
1	ALTO HOSPICIO	x	x
1	ANDACOLLO		x
1	ANTOFAGASTA	x	x
1	ARICA	x	x
1	CALAMA	x	x
1	COPIAPÓ	x	x
1	COQUIMBO	x	x
1	IQUIQUE	x	x
1	LA SERENA	x	x
1	MEJILLONES	x	
1	OVALLE	x	x
2	ALGARROBO	x	x
2	CABRERO	x	
2	CALERA	x	
2	CHIGUAYANTE	x	x
2	CHILLÁN	x	x
2	CONCEPCIÓN	x	x
2	CONCÓN	x	x
2	CORONEL	x	x
2	CURANILAHUE		x
2	CURICÓ	x	x
2	EL QUISCO		x
2	EL TABO		x
2	FLORIDA		x
2	HUALPÉN	x	x
2	HUALQUI		x
2	LIMACHE	x	
2	LOS ÁNGELES	x	
2	MACHALÍ		x
2	MAULE		x
2	MULCHÉN	x	
2	NOGALES	x	x
2	PENCO	x	x
2	QUILPUÉ	x	x
2	RANCAGUA	x	x
2	SAN ANTONIO		x
2	SAN CARLOS		x
2	SAN FELIPE	x	
2	SAN PEDRO DE LA PAZ	x	x
2	TALCA	x	x
2	TALCAHUANO	x	x
2	TOMÉ		x
2	TUCAPEL		x
2	VALPARAÍSO	x	x
2	VILLA ALEMANA	x	x
2	VIÑA DEL MAR	x	x

Macrozone	Municipalities	Refreshment 2024	Panel (after recontact)
3	ANCUD		x
3	ANGOL		x
3	AYSÉN	x	x
3	COYHAIQUE	x	x
3	FREIRE		x
3	NACIMIENTO		x
3	NATALES		x
3	PADRE LAS CASAS		x
3	PORVENIR	x	
3	PUERTO MONTT	x	x
3	PUERTO VARAS		x
3	PUNTA ARENAS	x	x
3	TEMUCO	x	x
3	VALDIVIA	x	x
4	BUIN	x	x
4	CERRILLOS	x	x
4	CERRO NAVIA	x	x
4	COLINA		x
4	CONCHALÍ	x	x
4	EL BOSQUE	x	x
4	ESTACIÓN CENTRAL	x	x
4	HUECHURABA	x	x
4	INDEPENDENCIA	x	x
4	ISLA DE MAIPO		x
4	LA CISTERNA	x	x
4	LA FLORIDA	x	x
4	LA GRANJA	x	x
4	LA PINTANA	x	x
4	LA REINA	x	x
4	LAMPA		x
4	LAS CONDES	x	x
4	LO BARNECHEA	x	x
4	LO ESPEJO	x	x
4	LO PRADO	x	x
4	MACUL	x	x
4	MAIPÚ	x	x
4	ÑUÑOA	x	x
4	PADRE HURTADO	x	x
4	PEDRO AGUIRRE CERDA	x	x
4	PEÑAFLOR		x
4	PEÑALOLÉN	x	x
4	PROVIDENCIA	x	x
4	PUDAHUEL	x	x
4	PUENTE ALTO	x	x
4	QUILICURA	x	x
4	QUINTA NORMAL	x	x
4	RECOLETA	x	x
4	RENCA	x	x
4	SAN BERNARDO	x	x
4	SAN JOAQUÍN	x	x
4	SAN MIGUEL	x	x
4	SAN RAMÓN	x	x
4	SANTIAGO	x	x
4	TALAGANTE		x
4	VITACURA	x	x

Annex 3: Rubin's Rules

Rubin's Rules are a set of statistical principles proposed by **Donald Rubin** for combining results from a **Multiple Imputation (MI)** process. These rules enable accurate and efficient parameter and error estimation after imputing multiple incomplete datasets.

Mean of an Imputed Variable

Suppose m imputed datasets have been generated. Then, a parameter of interest Q (e.g., a mean, regression coefficient, etc.) is calculated as:

$$\bar{Q} = \frac{1}{m}\sum_{i=1}^{m} Q_i$$

Where Q_i is the estimate of the parameter in the i-th imputed dataset. That is, the estimator (mean, median, regression coefficient, etc.) for an imputed variable is the average of the estimates obtained from each of the m datasets.

Variance of an Imputed Variable

The total variance of an imputed variable has two components:

1. **Within-imputation variance**:

$$W = \frac{1}{m}\sum_{i=1}^{m} V_i$$

   Where V_i is the variance of Q_i.

2. **Between-imputation variance**:

$$B = \frac{1}{m-1}\sum_{i=1}^{M} (Q_i - \bar{Q})^2$$

Thus, the **total variance** is:

$$T = W + \left(1 + \frac{1}{m}\right) \cdot B$$

Confidence Intervals

The confidence interval for an imputed variable using Rubin's Rules is:

$$\bar{Q} \pm t_{v,1-\frac{\alpha}{2}} \cdot \sqrt{T}$$

Where v is the adjusted degrees of freedom:

$$v = \frac{(m-1)\left(1 + \frac{W}{\left(1+\frac{1}{m}\right)B}\right)^2}{1 + \frac{1}{m+1}\left(\frac{W}{\left(1+\frac{1}{m}\right)B}\right)^2}$$

Regression and Statistical Significance

For each of the m imputed datasets, a regression model is estimated using **Ordinary Least Squares (OLS)**, retrieving the estimated coefficients β_i and their corresponding variances V_i (i.e., squared standard errors).

The **pooled regression coefficient** is:

$$\bar{\beta} = \frac{1}{m}\sum_{i=1}^{m} \beta_i$$

The **total variance** of these coefficients is:

$$T = \underbrace{\frac{1}{m}\sum_{i=1}^{m} V_i}_{W} + \left(1 + \frac{1}{m}\right) \cdot \underbrace{\frac{1}{m-1}\sum_{i=1}^{M} (\beta_i - \bar{\beta})^2}_{B}$$

The **standard error** is:

$$SE(\bar{\beta}) = \sqrt{T}$$

And the **t-statistic** is:

$$t = \frac{\bar{\beta}}{SE(\bar{\beta})}$$

This statistic is compared against a **t-distribution** $t_{v,1-\frac{\alpha}{2}}$, where v is the adjusted degrees of freedom defined above.

Annex 4: Missing Data Analysis – HFS 2024

Block	Feature	Description	Observations	No missings	Missings (N)	Missings (%)	Method
Block 1	yoprinm	Ingreso relacionado a la actividad principal del individuo	5841	4894	947	16%	MICE
	yosec	Ingreso relacionado otras ocupaciones distintas a la ocupación principal del individuo	787	553	234	30%	MICE
	yoprin_otro_1	OIAP: bonificaciones y gratificaciones	967	775	192	20%	MICE
	yoprin_otro_2	OIAP: remuneración en especies o regalías	107	84	23	21%	MICE
	yoprin_otro_3	OIAP: retiro de productos de su negocio (no agrícolas) para consumo propio	35	27	8	23%	HOTDECK
	yoprin_otro_4	OIAP: otros	27	24	3	11%	HOTDECK
	horas_trabajadas	Horas trabajadas	5841	5680	161	3%	MICE

Block	Feature	Description	Observations	No missings	Missings (N)	Missings (%)	Method
Block 2	y_pen_1	Pensión de alimentos	408	390	18	4%	HOTDECK
	y_pen_2	Dinero aportado por personas ajenas al hogar (residentes Chile)	185	177	8	4%	HOTDECK
	y_pen_3	Dinero aportado por personas ajenas al hogar (no residentes Chile)	31	29	2	6%	HOTDECK
	y_pen_6	Pensión de vejez, jubilación o rentas vitalicias	792	732	60	8%	HOTDECK
	y_pen_7	Pensión de invalidez	138	128	10	7%	HOTDECK
	y_pen_8	Montepío o pensión de viudez	170	154	16	9%	HOTDECK
	y_pen_9	Pensión de orfandad	18	15	3	17%	HOTDECK
	y_pen_10	Seguro de cesantía	65	51	14	22%	HOTDECK
	y_sub_4	Pensión básica solidaria de vejez o aporte previsional solidario	850	802	48	6%	HOTDECK
	y_sub_5	Pensión básica solidaria de invalidez	159	145	14	9%	HOTDECK
	y_sub_11	Subsidio familiar, duplo, discapacidad, cesantía	139	137	2	1%	HOTDECK
	y_sub_12	Subsidio al agua potable	373	315	58	16%	HOTDECK
	y_sub_13	Sistema de protección social"	20	16	4	20%	HOTDECK
	y_sub_14	Asignación familiar	273	260	13	5%	HOTDECK
	y_sub_15	Otros subsidios del estado	153	143	10	7%	HOTDECK
	y_otro_1	Intereses por ahorro	376	252	124	33%	HOTDECK
	y_otro_2	Dividendos por acciones	63	46	17	27%	HOTDECK
	y_otro_3	Retiro de utilidades de su negocio	53	42	11	21%	HOTDECK
	y_otro_4	Retiro de productos de su negocio	20	16	4	20%	HOTDECK
	y_otro_5	Productos huerto familiar	7	5	2	29%	HOTDECK
	y_otro_6	Venta de productos caseros	50	41	9	18%	HOTDECK
	y_otro_7	Finiquito o indemnización laboral	144	135	9	6%	HOTDECK
	y_otro_8	Devolución de impuestos	474	419	55	12%	HOTDECK
	y_otro_9	Donanciones de instituciones o personas ajenas al hogar	12	10	2	17%	HOTDECK
	y_otro_14	Arriendo de propiedades	280	270	10	4%	HOTDECK
	y_otro_15	Arriendo de maquinarias, herramientas o implementos	5	4	1	20%	HOTDECK
	y_otro_16	Arriendo de vehículos de transporte	16	13	3	19%	HOTDECK
	y_otro_18	Otros Ingresos	40	38	2	5%	HOTDECK

Block	Feature	Description	Observations	No missings	Missings (N)	Missings (%)	Method
Block 2	act_acc	Valor de acciones transadas en bolsa	119	82	37	31%	HOTDECK
	act_fm	Valor de fondos mutuos	386	308	78	20%	HOTDECK
	act_rv	Valor de otros activos financieros de renta variable	28	21	7	25%	HOTDECK
	act_ps	Valor de participación en sociedades o en fondos de inversión	113	73	40	35%	HOTDECK
	act_cripto	Valor de criptoactivos	78	67	11	14%	HOTDECK
	act_ca	Valor de cuentas de ahorro	779	627	152	20%	HOTDECK
	act_apv	Valor de ahorro previsional voluntario	368	256	112	30%	HOTDECK
	act_rf	Valor de instrumentos de renta fija	433	348	85	20%	HOTDECK
	act_seg	Valor de seguros con ahorro	206	131	75	36%	HOTDECK
	act_acc_12	Valor de acciones transadas en bolsa, inversión en últimos 12 meses	119	48	71	60%	HOTDECK
	act_fm_12	Valor de otros activos financieros de renta variable, inversión en últimos 12 meses	386	222	164	42%	HOTDECK
	act_rv_12	Valor de otros activos financieros de renta variable, inversión en últimos 12 meses	28	17	11	39%	HOTDECK
	act_ps_12	Valor de participación en sociedades o en fondos de inversión, inverisón en últimos 12 meses	113	41	72	64%	HOTDECK
	act_cripto_12	Valor de criptoactivos, inversión en últimos 12 meses	78	45	33	42%	HOTDECK
	act_ca_12	Valor de cuentas de ahorro, inversión en últimos 12 meses	779	632	147	19%	HOTDECK
	act_apv_12	Valor de ahorro previsional voluntario, inversión en últimos 12 meses	368	201	167	45%	HOTDECK
	act_rf_12	Valor de instrumentos de renta fija, inversión en últimos 12 meses	433	327	106	24%	HOTDECK
	act_seg_12	Valor de seguros con ahorro, inversión en últimos 12 meses	206	119	87	42%	HOTDECK
	act_aut	Valor de autos o camionetas (incluye taxis)	2648	2514	134	5%	HOTDECK
	act_mot	Valor de motos	208	198	10	5%	HOTDECK
	act_utt	Valor de furgones y utilitarios	84	78	6	7%	HOTDECK
	act_otrosvehic	Valor de otros activos automotrices (ej: embarcaciones, aviones, helicópteros)	16	15	1	6%	HOTDECK
	act_vp	Valor de vivienda propia	2607	2120	487	19%	HOTDECK
	act_otp_1	Valor de otras propiedades	748	664	84	11%	HOTDECK
	act_otp_2	Valor de otras propiedades	228	208	20	9%	HOTDECK
	act_otp_3	Valor de otras propiedades	70	63	7	10%	HOTDECK
	act_ahcta	Valor de ahorro en cuenta corriente/cuenta vista	1181	968	213	18%	HOTDECK
	habah_m	Monto ahorrado en los últimos 12 meses	1573	1344	229	15%	HOTDECK
	cap_pen	Monto ahorrado en los últimos 12 meses	1313	1173	140	11%	HOTDECK
	act_otros	Valor de otros activos reales (maquinaria industrial/agrícola, cabezas de ganado u otros animales, obras de arte, etc.)	53	45	8	15%	HOTDECK

Block	Feature	Description	Observations	No missings	Missings (N)	Missings (%)	Method
Block 2	m_vp_compra	Valor de compra de vivienda principal	2607	1718	889	34%	HOTDECK
	m_op_venta_1	Valor de venta de la otra propiedad 1	748	664	84	11%	HOTDECK
	m_op_venta_2	Valor de venta de la otra propiedad 2	228	208	20	9%	HOTDECK
	m_op_venta_3	Valor de venta de la otra propiedad 3	70	63	7	10%	HOTDECK
	m_vp_fr	Monto financiado de vivienda principal con recursos propios	1883	1384	499	27%	HOTDECK
	m_vp_fs	Monto financiado de vivienda principal con subsidio	862	453	409	47%	HOTDECK
	m_vp_fch	Monto financiado de vivienda principal con crédito hipotecario	1200	904	296	25%	HOTDECK
	m_vp_foc	Monto financiado de vivienda principal con otras fuentes	129	96	33	26%	HOTDECK
	m_vp_ftr	Monto financiado de vivienda principal transferencias	142	52	90	63%	HOTDECK
	m_vp_ffa	Monto financiado de vivienda principal a través de créditos de familiares o amigos	75	51	24	32%	HOTDECK
	m_otp_fr_1	Monto financiado de otra propiedad 1 con recursos propios	504	444	60	12%	HOTDECK
	m_otp_fs_1	Monto financiado de otra propiedad 1 con subsidio	114	40	74	65%	HOTDECK
	m_otp_fch_1	Monto financiado de otra propiedad 1 con crédito hipotecario	344	307	37	11%	HOTDECK
	m_otp_foc_1	Monto financiado de otra propiedad 1 con otros créditos de instituciones financieras	42	38	4	10%	HOTDECK
	m_otp_ftr_1	Monto financiado de otra propiedad 1 con transferencias	99	84	15	15%	HOTDECK
	m_otp_ffa_1	Monto financiado de otra propiedad 1 con créditos de familiares o amigos	22	18	4	18%	HOTDECK
	m_otp_fr_2	Monto financiado de otra propiedad 2 con recursos propios	164	148	16	10%	HOTDECK
	m_otp_fs_2	Monto financiado de otra propiedad 2 con subsidio	17	7	10	59%	HOTDECK
	m_otp_fch_2	Monto financiado de otra propiedad 2 con crédito hipotecario	92	79	13	14%	HOTDECK
	m_otp_foc_2	Monto financiado de otra propiedad 2 con otros créditos de instituciones financieras	12	11	1	8%	HOTDECK
	m_otp_ftr_2	Monto financiado de otra propiedad 2 con transferencias	33	31	2	6%	HOTDECK
	m_otp_fr_3	Monto financiado de otra propiedad 3 con recursos propios	56	47	9	16%	HOTDECK
	m_otp_fs_3	Monto financiado de otra propiedad 3 con subsidio	7	2	5	71%	HOTDECK
	m_otp_fch_3	Monto financiado de otra propiedad 3 con crédito hipotecario	26	24	2	8%	HOTDECK
	m_otp_foc_3	Monto financiado de otra propiedad 3 con otros créditos de instituciones financieras	4	3	1	25%	HOTDECK

Block	Feature	Description	Observations	No missings	Missings (N)	Missings (%)	Method
Block 2	div_chvp	Dividendo del crédito hipotecario para vivienda principal	686	662	24	3%	HOTDECK
	div_ocvp	Dividendo del crédito complementario para vivienda principal	28	23	5	18%	HOTDECK
	div_favp	Dividendo del crédito con familiares y amigos para vivienda principal	13	8	5	38%	HOTDECK
	div_chotp_1	Dividendo del crédito hipotecario para otras propiedades	231	223	8	3%	HOTDECK
	div_chotp_2	Dividendo del crédito hipotecario para otras propiedades	53	51	2	4%	HOTDECK
	div_chotp_3	Dividendo del crédito hipotecario para otras propiedades	16	15	1	6%	HOTDECK
	div_ocotp_2	Dividendo del crédito complementario para otra propiedad 2	4	3	1	25%	HOTDECK
	div_faotp_1	Dividendo del crédito con familiares y amigos para otra propiedad 1	5	3	2	40%	HOTDECK
	pp_m_chvp	Plazo por pagar en meses del crédito hipotecario para vivienda principal	686	683	3	0%	HOTDECK
	pp_m_ocvp	Plazo por pagar en meses del crédito complementario para vivienda principal	28	26	2	7%	HOTDECK
	pp_m_favp	Plazo por pagar en meses del crédito con familiares y amigos para vivienda principal	9	7	2	22%	HOTDECK
	pp_m_chotp_1	Plazo por pagar en meses del crédito hipotecario para otras propiedades	231	230	1	0%	HOTDECK
	pp_m_faotp_1	Plazo por pagar en meses del crédito con familiares y amigos para otra propiedad 1	5	3	2	40%	HOTDECK
	pp_m_chotp_2	Plazo por pagar en meses del crédito hipotecario para otras propiedades	53	52	1	2%	HOTDECK
	pp_m_faotp_2	Plazo por pagar en meses del crédito con familiares y amigos para otra propiedad 2	2	1	1	50%	HOTDECK

Block	Feature	Description	Observations	No missings	Missings (N)	Missings (%)	Method
Block 2	gasto_m_super	Gasto mensual en Supermercado	4649	4467	182	4%	HOTDECK
	gasto_m_almacen	Gasto mensual en Ferias libres, almacenes y negocios de barrio	4649	4393	256	6%	HOTDECK
	gasto_m_agua	Gasto mensual en agua potable	4649	4467	182	4%	HOTDECK
	gasto_m_luz	Gasto mensual en luz o electricidad	4649	4464	185	4%	HOTDECK
	gasto_m_gas	Gasto mensual en gas (licuado o por cañeria)	4649	4421	228	5%	HOTDECK
	gasto_m_lena	Gasto mensual en leña	4649	4216	433	9%	HOTDECK
	gasto_m_internet	Gasto mensual en servicios de internet y/o televisión pagada (cable, satelital o servicios de streaming)	4649	4323	326	7%	HOTDECK
	gasto_m_fono	Gasto mensual en telefónia fija o móvil	4649	4313	336	7%	HOTDECK
	gasto_m_gcomun	Gasto mensual en gastos comunes de condominio o edificio	4649	4147	502	11%	HOTDECK
	gasto_m_tpublico	Gasto mensual en transporte público (buses, micros, metro, colectivos, trenes, taxis y taxis en aplicaciones)	4649	4327	322	7%	HOTDECK
	gasto_m_bencina	Gasto mensual en bencina o combustibles para vehículos del hogar	4649	4268	381	8%	HOTDECK
	gasto_m_tag	Gasto mensual en peajes o tag	4649	4070	579	12%	HOTDECK
	gasto_m_cmedicas	Gastos en consultas médicas (incluya dentales, psicológicas, kinesiológicas, etc.)	4649	4277	372	8%	HOTDECK
	gasto_m_exmedicos	Gastos en exámenes médicos y procedimientos médicos ambulatorios	4649	4236	413	9%	HOTDECK
	gasto_m_medicamentos	Gasto en medicamentos	4649	4289	360	8%	HOTDECK
	gasto_m_electro	Gasto en electrodomésticos pequeños	4649	4558	91	2%	HOTDECK
	gasto_m_mantenviv	Gasto en servicios y/o compra de materiales para conservación de la vivienda	4649	4566	83	2%	HOTDECK
	gasto_m_vestuario	Gasto en vestuario y zapatos	4649	4555	94	2%	HOTDECK
	gasto_m_celulares	Gasto en compra de equipos de telefonía móvil	4649	4572	77	2%	HOTDECK
	gasto_m_elecgrandes	Gasto en electrodomésticos grandes	4649	4582	67	1%	HOTDECK
	gasto_m_muebles	Gasto en muebles: mensualizado	4649	4588	61	1%	HOTDECK
	m_arriendo_vp	Monto de arriendo de la vivienda	1233	1201	32	3%	HOTDECK
	yai	Alquiler imputado vivienda propia	3408	3099	309	9%	HOTDECK

Block	Feature	Description	Observations	No missings	Missings (N)	Missings (%)	Method
Block 3	cf_td1_1	Carga financiera en deudas asociadas a tarjetas de crédito bancarias	796	718	78	10%	HOTDECK
	cf_td1_2	Carga financiera en deudas asociadas a tarjetas de crédito bancarias	311	274	37	12%	HOTDECK
	cf_td1_3	Carga financiera en deudas asociadas a tarjetas de crédito bancarias	91	70	21	23%	HOTDECK
	cf_td2_1	Carga financiera en deudas asociadas a línea de crédito	282	229	53	19%	HOTDECK
	cf_td2_2	Carga financiera en deudas asociadas a línea de crédito	76	55	21	28%	HOTDECK
	cf_td2_3	Carga financiera en deudas asociadas a línea de crédito	20	12	8	40%	HOTDECK
	cf_td3_1	Carga financiera en deudas asociadas a tarjetas de crédito de casas comerciales	914	811	103	11%	HOTDECK
	cf_td3_2	Carga financiera en deudas asociadas a tarjetas de crédito de casas comerciales	342	286	56	16%	HOTDECK
	cf_td3_3	Carga financiera en deudas asociadas a tarjetas de crédito de casas comerciales	113	83	30	27%	HOTDECK
	cf_td4_1	Carga financiera en deudas asociadas a préstamos de consumo en casas comerciales	154	149	5	3%	HOTDECK
	cf_td4_2	Carga financiera en deudas asociadas a préstamos de consumo en casas comerciales	28	26	2	7%	HOTDECK
	cf_td4_3	Carga financiera en deudas asociadas a préstamos de consumo en casas comerciales	8	5	3	38%	HOTDECK
	cf_td5_1	Carga financiera en deudas asociadas a préstamos de consumo bancarios o en financieras	480	455	25	5%	HOTDECK
	cf_td5_2	Carga financiera en deudas asociadas a préstamos de consumo bancarios o en financieras	105	89	16	15%	HOTDECK
	cf_td5_3	Carga financiera en deudas asociadas a préstamos de consumo bancarios o en financieras	21	18	3	14%	HOTDECK
	cf_td6_1	Carga financiera en deudas asociadas a préstamos de cajas de compensación, cooperativas u otros similares	383	353	30	8%	HOTDECK
	cf_td6_3	Carga financiera en deudas asociadas a préstamos de cajas de compensación, cooperativas u otros similares	7	6	1	14%	HOTDECK
	cf_td7_1	Carga financiera en deudas asociadas a créditos automotrices	185	177	8	4%	HOTDECK
	cf_td8_1	Carga financiera en deudas educacionales	337	271	66	20%	HOTDECK
	cf_td8_2	Carga financiera en deudas educacionales	70	56	14	20%	HOTDECK
	cf_td9_1	Carga Financiera Efectiva(monto facturado) de Préstamos parientes o amigos	92	90	2	2%	HOTDECK
	cf_td12_1	Carga Financiera Efectiva(monto facturado) de Fiado	26	25	1	4%	HOTDECK
	cf_td13_1	Carga Financiera Efectiva(monto facturado) de Otras fuentes de crédito	41	39	2	5%	HOTDECK
	ct_td1_1	Cupo total de la tarjeta de crédito 1	813	740	73	9%	HOTDECK
	ct_td1_2	Cupo total de la tarjeta de crédito 2	328	290	38	12%	HOTDECK
	ct_td1_3	Cupo total de la tarjeta de crédito 3	108	76	32	30%	HOTDECK
	ct_td2_1	Cupo total de la línea de crédito 1	305	250	55	18%	HOTDECK
	ct_td2_2	Cupo total de la línea de crédito 2	99	59	40	40%	HOTDECK
	ct_td2_3	Cupo total de la línea de crédito 3	43	14	29	67%	HOTDECK
	ct_td3_1	Cupo total de la tarjeta de casas comerciales 1	935	829	106	11%	HOTDECK
	ct_td3_2	Cupo total de la tarjeta de casas comerciales 2	363	308	55	15%	HOTDECK
	ct_td3_3	Cupo total de la tarjeta de casas comerciales 3	134	103	31	23%	HOTDECK

Block	Feature	Description	Observations	No missings	Missings (N)	Missings (%)	Method
Block 3	m_td1_1	Monto inicial de la deuda asociada a tarjetas de crédito bancarias	796	759	37	5%	HOTDECK
	m_td1_2	Monto inicial de la deuda asociada a tarjetas de crédito bancarias	311	292	19	6%	HOTDECK
	m_td1_3	Monto inicial de la deuda asociada a tarjetas de crédito bancarias	91	77	14	15%	HOTDECK
	m_td2_1	Monto inicial de la deuda asociada a línea de crédito	282	240	42	15%	HOTDECK
	m_td2_2	Monto inicial de la deuda asociada a línea de crédito	76	58	18	24%	HOTDECK
	m_td2_3	Monto inicial de la deuda asociada a línea de crédito	20	13	7	35%	HOTDECK
	m_td3_1	Monto inicial de la deuda asociada a tarjetas de crédito de casas comerciales	914	849	65	7%	HOTDECK
	m_td3_2	Monto inicial de la deuda asociada a tarjetas de crédito de casas comerciales	342	310	32	9%	HOTDECK
	m_td3_3	Monto inicial de la deuda asociada a tarjetas de crédito de casas comerciales	113	99	14	12%	HOTDECK
	m_td4_1	Monto inicial de la deuda asociada a préstamos de consumo en casas comerciales	154	146	8	5%	HOTDECK
	m_td4_2	Monto inicial de la deuda asociada a préstamos de consumo en casas comerciales	28	27	1	4%	HOTDECK
	m_td4_3	Monto inicial de la deuda asociada a préstamos de consumo en casas comerciales	8	7	1	13%	HOTDECK
	m_td5_1	Monto inicial de la deuda asociada a préstamos de consumo bancarios o en financieras	480	452	28	6%	HOTDECK
	m_td5_2	Monto inicial de la deuda asociada a préstamos de consumo bancarios o en financieras	105	92	13	12%	HOTDECK
	m_td5_3	Monto inicial de la deuda asociada a préstamos de consumo bancarios o en financieras	21	18	3	14%	HOTDECK
	m_td6_1	Monto inicial de la deuda asociada a préstamos de cajas de compensación, cooperativas u otros similares	383	360	23	6%	HOTDECK
	m_td6_2	Monto inicial de la deuda asociada a préstamos de cajas de compensación, cooperativas u otros similares	53	51	2	4%	HOTDECK
	m_td7_1	Monto inicial de la deuda asociada a créditos automotrices	185	169	16	9%	HOTDECK
	m_td8_1	Monto inicial de la deuda asociada a créditos educacionales	460	350	110	24%	HOTDECK
	m_td8_2	Monto inicial de la deuda asociada a créditos educacionales	93	61	32	34%	HOTDECK
	m_td8_3	Monto inicial de la deuda asociada a créditos educacionales	8	7	1	13%	HOTDECK
	m_td9_1	Monto inicial de la deuda asociada a préstamos de parientes o amigos	154	144	10	6%	HOTDECK
	m_td12_1	Monto inicial de la deuda asociada a pedir fiado	50	49	1	2%	HOTDECK
	m_td13_1	Monto inicial de la deuda asociada a créditos de otras fuentes	51	46	5	10%	HOTDECK

Block	Feature	Description	Observations	No missings	Missings (N)	Missings (%)	Method
Block 3	pp_td1_1	Plazo por pagar en meses en deudas asociadas a tarjetas de crédito bancarias	796	756	40	5%	HOTDECK
	pp_td1_2	Plazo por pagar en meses en deudas asociadas a tarjetas de crédito bancarias	311	295	16	5%	HOTDECK
	pp_td1_3	Plazo por pagar en meses en deudas asociadas a tarjetas de crédito bancarias	91	78	13	14%	HOTDECK
	pp_td2_1	Plazo por pagar en meses en deudas asociadas a líneas de crédito	282	256	26	9%	HOTDECK
	pp_td2_2	Plazo por pagar en meses en deudas asociadas a líneas de crédito	76	62	14	18%	HOTDECK
	pp_td2_3	Plazo por pagar en meses en deudas asociadas a líneas de crédito	20	14	6	30%	HOTDECK
	pp_td3_1	Plazo por pagar en meses en deudas asociadas a tarjetas de crédito de casas comerciales	914	841	73	8%	HOTDECK
	pp_td3_2	Plazo por pagar en meses en deudas asociadas a tarjetas de crédito de casas comerciales	342	301	41	12%	HOTDECK
	pp_td3_3	Plazo por pagar en meses en deudas asociadas a tarjetas de crédito de casas comerciales	113	95	18	16%	HOTDECK
	pp_td4_1	Plazo por pagar en meses en deudas asociadas a préstamos de consumo en casas comerciales	154	141	13	8%	HOTDECK
	pp_td4_2	Plazo por pagar en meses en deudas asociadas a préstamos de consumo en casas comerciales	28	26	2	7%	HOTDECK
	pp_td4_3	Plazo por pagar en meses en deudas asociadas a préstamos de consumo en casas comerciales	8	5	3	38%	HOTDECK
	pp_td5_1	Plazo por pagar en meses en deudas asociadas a préstamos de consumo bancarios o en financieras	480	449	31	6%	HOTDECK
	pp_td5_2	Plazo por pagar en meses en deudas asociadas a préstamos de consumo bancarios o en financieras	105	91	14	13%	HOTDECK
	pp_td5_3	Plazo por pagar en meses en deudas asociadas a préstamos de consumo bancarios o en financieras	21	18	3	14%	HOTDECK
	pp_td6_1	Plazo por pagar en meses en deudas asociadas a créditos automotrices	383	356	27	7%	HOTDECK
	pp_td6_2	Plazo por pagar en meses en deudas asociadas a créditos automotrices	53	50	3	6%	HOTDECK
	pp_td7_1	Plazo por pagar en meses en deudas asociadas a créditos de cajas de compensación, cooperativas u otros similares	185	175	10	5%	HOTDECK
	pp_td8_1	Plazo por pagar en meses en deudas asociadas a créditos educacionales	337	189	148	44%	HOTDECK
	pp_td8_2	Plazo por pagar en meses en deudas asociadas a créditos educacionales	70	37	33	47%	HOTDECK
	pp_td8_3	Plazo por pagar en meses en deudas asociadas a créditos educacionales	4	2	2	50%	HOTDECK

Block	Feature	Description	Observations	No missings	Missings (N)	Missings (%)	Method
Block 3	pr_td1_1	Pago total realizado en deudas asociadas a tarjetas de crédito bancarias	796	709	87	11%	HOTDECK
	pr_td1_2	Pago total realizado en deudas asociadas a tarjetas de crédito bancarias	311	273	38	12%	HOTDECK
	pr_td1_3	Pago total realizado en deudas asociadas a tarjetas de crédito bancarias	91	69	22	24%	HOTDECK
	pr_td3_1	Pago total realizado en deudas asociadas a tarjetas de crédito de casas comerciales	914	765	149	16%	HOTDECK
	pr_td3_2	Pago total realizado en deudas asociadas a tarjetas de crédito de casas comerciales	342	260	82	24%	HOTDECK
	pr_td3_3	Pago total realizado en deudas asociadas a tarjetas de crédito de casas comerciales	113	79	34	30%	HOTDECK
	pr_td4_1	Pago total realizado en deudas asociadas a préstamos de consumo en casas comerciales	154	135	19	12%	HOTDECK
	pr_td4_2	Pago total realizado en deudas asociadas a préstamos de consumo en casas comerciales	28	25	3	11%	HOTDECK
	pr_td4_3	Pago total realizado en deudas asociadas a préstamos de consumo en casas comerciales	8	6	2	25%	HOTDECK
	pr_td5_1	Pago total realizado en deudas asociadas a préstamos de consumo bancarios o en financieras	479	434	45	9%	HOTDECK
	pr_td5_2	Pago total realizado en deudas asociadas a préstamos de consumo bancarios o en financieras	105	86	19	18%	HOTDECK
	pr_td5_3	Pago total realizado en deudas asociadas a préstamos de consumo bancarios o en financieras	21	18	3	14%	HOTDECK
	pr_td6_1	Pago total realizado en deudas asociadas a créditos de cajas de compensación, cooperativas u otros similares	383	343	40	10%	HOTDECK
	pr_td6_2	Pago total realizado en deudas asociadas a créditos de cajas de compensación, cooperativas u otros similares	53	51	2	4%	HOTDECK
	pr_td6_3	Pago total realizado en deudas asociadas a créditos de cajas de compensación, cooperativas u otros similares	7	6	1	14%	HOTDECK
	pr_td7_1	Pago total realizado en deudas asociadas a créditos automotrices	185	170	15	8%	HOTDECK
	pr_td7_2	Pago total realizado en deudas asociadas a créditos automotrices	16	14	2	13%	HOTDECK

Annex 5: Box-Cox Transformation

Box-Cox transformation is a technique used to stabilize variance and approximate normality in the distribution of a variable. This is particularly useful in the context of **multiple imputations**, where the models used tend to perform better under the assumption of normality.

The main objectives of applying this transformation are:

- To reduce **distribution skewness**
- To improve **fit to normality**
- To facilitate the application of **multiple imputation methods**

The **Box-Cox transformation** is defined as:

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda}, & if\ \lambda \neq 0 \\ \log(y), & if\ \lambda = 0 \end{cases}$$

Where:

y is the original variable
λ is the transformation parameter.

In the specific case of the HFS, the `bcskew0` command in Stata is used to apply Box-Cox transformations with the specific goal of reducing the skewness of a variable. This command searches for the transformation parameter λ that minimizes the skewness of the transformed variable. Unlike the traditional Box-Cox transformation, which typically estimates λ via maximum likelihood to approximate normality, `bcskew0` focuses on finding the value of λ that brings skewness as close to zero as possible.