Methodological Document

Supplementary Survey of Income (ESI 2021)

National Statistics Institute

July 2022

DEPARTMENT OF LABOUR STATISTICS
DEPARTMENT OF OPERATIONS FOR SOCIAL STATISTICS

National Statistics Institute

Methodological Document, Supplementary Survey of Income, ESI 2021

Technical Subdirector
Leonardo González Allendes

Operations Subdirector
Daniela Moraga Farias

Head of the Department of Labour Statistics
David Niculcar Castillo

Head of the Department of Operations for Social Statistics
Rodrigo Machuca Méndez

National Directorate: Morandé No. 801, piso 22, Santiago, Chile.
Telephone: 232461010
Fax: 562 267 12 169
Postal Code: 8340148
Website: www.ine.cl
E-mail: ine@ine.cl
Santiago, Chile

Table of contents

I.   Introduction

The National Statistics Institute (INE) of Chile is the agency in charge of producing and disseminating official statistics in the country. These statistics are valuable because they provide us with information about the conditions of the country's economy, population, and territory, and the information assists the State in implementing initiatives that benefit the entire population.

INE's mission states that in addition to producing statistics, it must guarantee their technical quality and ensure the delivery of information in a reliable, timely, relevant, comparable, and easily accessible manner. For these purposes, the development of standardized documentation and metadata plays an important role.

Among INE's statistical products is the Supplementary Survey of Income (*Encuesta Suplementaria de Ingresos*, or ESI), whose objective[1] is to describe the labour income of people who are classified as employed in the National Employment Survey (*Encuesta Nacional de Empleo,* or ENE) as well as income from activities other than the main job, both at the national and regional levels. As a statistical product collected by INE, the ESI must meet INE's objectives and guidelines: it must provide access to the results, facilitate their understanding, and ensure clarity and transparency, all with the assistance of appropriate metadata.

The purpose of this document is to describe the processes and methodologies used in the production of the ESI in order to ensure that those who access it have comprehensive information on the statistical products, from construction to the delivery of final results, thus increasing understanding of the survey and maximizing the usefulness of the information provided.

The following chapters provide information on the background of the survey, the reference framework, the statistical design, the methodologies of data collection and processing, and the methodology of survey analysis. Other topics include innovations and questionnaire refinements made in 2010, the most important milestones and major changes from 2010 to 2021, the selection of the sample, the presentation of results, and the delivery of metadata.

---

[1] Since the 2021 ESI, the objectives have focused on capturing and measuring labour income rather than on other sources of household income, which means that estimates of income at the household level will no longer be disseminated. This decision and its rationale were reported at the June 2021 meeting of the Expert Group on Labour Statistics. For more information, see the related document (in Spanish) at https://www.ine.gob.cl/docs/default-source/ocupacion-y-desocupacion/comites-y-notas-tecnicas/2021/comit%C3%A9-t%C3%A9cnico-de-expertos-estad%C3%ADsticas-laborales-(junio-2021).pdf?sfvrsn=355552fe_4.

II.   Context

2.1 Background of the survey

2.1.1   Context description

The Supplementary Survey of Income[2] (ESI) is a complementary module that is added to the current National Employment Survey (ENE). The ESI is conducted every year in the October–December quarter, and its purpose is to characterize the income of the main job of persons classified as employed in the ENE, as well as the income of secondary occupations distinct from the main job, both at the national and regional levels.

The ESI consists of a question module linked to the standard form of the employment survey. Thus, it constitutes a complementary source of information that captures information on labour income of the employed population. It is the only annual survey on labour income that is representative of all the regions of Chile.

The survey measures income from employment during the reference period under the concept of **current income** (GITCN, 2009).

Income from labour is net income (i.e., it excludes legal deductions and taxes), and it distinguishes between income from paid and self-employment jobs.

Since 2010, the ESI results come from a new questionnaire, which includes a series of improvements to the income questions that increase their precision. In addition, they were the first income measurement that was based on the current ENE, and thus the ESI was the first to incorporate the ENE's conceptual changes. For this reason, the results from 2010 onwards are not comparable with former versions.

2.1.2   Conceptual innovations and refinement of the income questionnaire that were carried out in 2010

▪   **Conceptual innovations**

As a module of the current ENE, the conceptual innovations in the measurement of employment and unemployment, which were incorporated in 2010, indirectly affected the ESI. These innovations are summarized in the following points:

1.1.   Updating the concept of employed person, which includes the explicit criterion of having worked at least one hour during the reference week of

---

[2] In 2015, the Supplementary Survey of Income returned to the use of its original name (from *Nueva Encuesta Suplementaria de Ingresos* to *Encuesta Suplementaria de Ingresos*). Throughout the document, we will refer to the surveys prior to 2010 as the "former ESI" and the "former ENE".

the survey and of having received income for the work (with the exception of unpaid family members).

1.2. Refining the filters for capturing employment status resulted in a more rigorous composition of the sub-populations of the ICSE[3], which, in turn, was of fundamental importance in the income module and in determining the starting question for members of the household.

1.3. Updating nomenclatures associated with level of education (ISCED-97), branch of economic activity (ISIC-Rev.3), and occupational group (ISCO-88)[4].

In addition, the question on branch of economic activity changed from asking about the establishment where the employed person works to asking about the enterprise where they work. Because an enterprise can include more than one establishment, the change can greatly affect the classification of a worker's economic activity.

As in previous versions, labour income always refers to net income. For paid employment jobs, legal discounts (social security, health, unemployment insurance, and the so-called "single tax"[5]) must be deducted from gross income. For self-employment jobs, the total costs associated with the activity performed must be subtracted from the gross income to obtain the net profit[6].

In a strict sense, and in accordance with international recommendations regarding income surveys, the disposable income measured by the ESI is the current income that can be used entirely for consumption or savings (investment) (United Nations, 2011). This concept was not altered with respect to the former measurement.

▪ **Refining the income questionnaire**

Improvements to the income questionnaire consisted of the following:

1.1. Ensuring that income from employment was fully consistent with the labour activity reported in the core module of the questionnaire.

---

[3] International Classification of Status in Employment (ICSE).

[4] These are the International Standard Classification of Education (ISCED), the International Standard Industrial Classification (ISIC), and the International Standard Classification of Occupations (ISCO).

[5] For honoraria workers, these payments amount to 11.5% of gross income, even if there is a subsequent reimbursement (effective until 31 December 2021).

[6] For the self-employed who self-assign a salary, this amount is added to the net benefits they report because it is part of their disposable income as natural persons.

This improvement to the form was the most important. Its objective is to distinguish whether the person's current job, that is, the one reported in the reference week, was the same as the one held in the month prior to the month of the survey.

This change meant establishing a clear link between the characterization of employment (provided by the central form of the survey) and information on income from employment (provided by the ESI). In the previous editions of the survey, it was not possible to verify this link, because nothing ensured that the job reported in the employment survey was the same as the one whose income was reported in the former ESI[7].

Thanks to this change, the analysis of employment information has been expanded to include information related to labour income.

1.2.   More precisely measuring the variable components and compensation of the remuneration of paid employed persons, establishing a clear distinction between both sources of income.

1.3.   Measuring the actual time worked of paid employed persons in their current job during the calendar month preceding the reference week. This improvement is important because it enables a direct link to be made between the hours worked and the income received for the work.

1.4.   Regarding the income of self-employed persons, the survey continues asking about net income, but it explains that those who self-assign a salary should add it to net income. Likewise, the question on own consumption was improved, and a question specifically associated with the use of one's own dwelling in the production process was included.

1.5.   Incorporating a question on income received during the reference month from jobs other than the current one that classified persons as employed and non-employed in the reference week. For the employed population, this kind of labour income conceptually refers to income from a secondary activity or and activity other than the current one, while in the case of the non-employed, inactive, or unemployed population, the measurement of labour income needs to consider that the activity status in the previous month may be different from the current situation, for example, because the person was laid off last month and is currently looking for a job. In short, this

---

[7] It should be noted that the reference periods of the surveys are different: the current ENE measures the previous week, while the ESI measures the calendar month prior to the reference week of the employment survey.

question provides indicators on recent non-employment, thus obtaining the entirety of income from all labour, known as income from employment[8].

1.6. Questions on income from other sources were also refined. Real estate and movable property rents were distinguished by determining whether they were of an agricultural nature. Likewise, the question on imputed rent was refined to distinguish the different types of home ownership and to capture relevant information for the subsequent imputation process of this variable. Regarding savings instruments, the possible interest collection brackets were updated and the reporting of non-response was improved.

1.7. To account for transfers, questions were expanded to capture the range of possible pensions, subsidies, and donations from which a person or household can benefit (pensions, allowances, special subsidies, scholarships, transfers between households, and other income from abroad). As a result, data collection improved both in terms of quality and coverage, resulting in higher levels than those recorded in the previous measurement.

1.8. Another major improvement to the income questionnaire was the emphasis on distinguishing the different scopes of non-response. Thus, "not sure" was differentiated from "no response" in the key questions of the module, and it is expected that this distinction will be further developed in subsequent versions of the survey[9].

- **Implications for the comparability of data**

In a strict sense, the ESI is not comparable with editions of the survey prior to 2010, which is due to the following factors:

2.1. The refinement of the questionnaire entailed a higher level of specificity and precision in capturing all the items that make up personal and household income. Therefore, although the income concepts are the same, the effect associated with the questionnaire means that the data cannot be compared.

---

[8] Income from employment is divided into income from wages and salaries, income of employers and own-account workers, and income from other work. This last item includes income from jobs other than the current one in the previous month (secondary activities) and labour income received by the inactive or unemployed persons in the month previous to the survey.

[9] For more information, see section 2.1.3, "Milestones and major changes, 2010–2021".

2.2. The new income module is an extension of the new employment form. Therefore, the concept of employment is not the same as in the former ESI[10].

2.3. In addition, the employed population was cleansed in such a way that only those persons who actually maintained their job in the two reference periods involved (reference week and reference month) are counted.

2.4. After updating the nomenclatures for branch of activity (ISIC Rev. 3), occupational group (ISCO 88), and level of education (ISCED 97), the aggregations of information no longer coincided with those of previous surveys.

Regarding total household income and its distribution, no conceptual innovations were implemented with respect to previous years; however, this does not guarantee its comparability, because the new filters of the current ENE resulted in a restructuring of the labour supply within households in which the employed population gained relative weight within the total population.

2.1.3 Milestones and major changes 2010–2021

▪ **Improvements to the questionnaire**

During ESI collection, experience in the field and improvements to operational and technical processes have led to some clarifications to the questionnaire, which are summarized below:

Table 1. Summary of modifications to the ESI questionnaire, 2010–2021.

Year of collection	Modifications
2011	- To facilitate the completion of surveys in the field, "not sure" and "no response" were included as options to a set of questions. - As a control mechanism to ensure the recording of employed persons in the previous month, filter questions on the income of paid employed and of self-employed persons were created (D1_opcion y D5_opcion). - Improvement in the completeness of the response options: "Do not receive" in D2, and option "nursery/kindergarten" in D3, while in D14

---

[10] In addition to being employed, the person must receive income from employment (except in the case of unpaid family workers). In addition, it is more rigorous in requiring at least one hour of work in the reference week of the survey.

Year of collection	Modifications
	"pensions from the State" was changed to "basic solidarity pension from the State". - Reformulation of some questions according to field experience and quality of information obtained: D6. - Merging of questions to facilitate capture: D12 and D13.
2012	- To facilitate sequencing of the questionnaire in the field, question D2 was moved while connectors were added to facilitate the survey. - The use of "not sure" and "no response" options for amounts, percentages, and brackets in questions D2, D3, and D13 was extended. - Expanded response options according to activity status in D9. - The variable income option "other type of incentives" in D2 was eliminated.
2015	- To improve information on the income of paid employed persons (question D2), a question on the frequency of reported income was added. - To improve information on the income of self-employed persons, a question was added on how much time was needed to earn the reported income (question D5). - The layout of question D12 was altered to improve its capture. - To improve the capture of income from financial assets (question D13), the number of brackets on interest was expanded. - To improve the capture of income from scholarships, question D14_11 was divided into two alternatives: monetary and non-monetary scholarships.
2016	- Given the low frequency and incidence of some of the alternatives in question D3, INE eliminated the questions on benefits or allowances for clothing and footwear, personal telephone, parking service, and per diems. However, these options can still be reported under "*other (specify)*". - Two sections for "calculations" were incorporated to facilitate the collection of complex information by the interviewer. - In order to improve capture, the layout of question D12 was modified. Specifically, a skip was added to the alternative "own, but pay mortgage", which goes directly to the estimate of home value, and it was reordered to facilitate the reading and skipping of the question.
2017	- Improvements were made in the wording of questions and instructions for the following questions: D1b, D2a, D2b, D2c, D3, D4, D5b, D6, D8, D9a, D11, D13 - The term "*viáticos*" (per diems) was eliminated from question D3. This decision is consistent with the elimination of the question on this benefit asked in 2016 (low prevalence).

Year of collection	Modifications
	- In D14, "NS/NR" (not sure, no response) was incorporated as an option for those who declare that they received an item of income, but did not indicate the amount.
2018	- In question D3, the definition of benefit is expanded to include anything that is "NOT necessary for the execution of their work". - In D12, option 5 "provided by the employer" and question 5.1 concerning anything "necessary for the execution of their work" were added, linking the type of dwelling (provided or necessary) with the benefit they may receive from the employer (link with question D3).
2019	- No changes.
2020	- No changes.
2021	- Questions D10 to D14, which concern income from other sources (property income, financial assets, and transfers), were eliminated.

- **Revision of the income module**

Until 2012, the estimated number of employed persons in the October-November-December quarter (OND) of the ENE and the ESI differed in only two aspects: estimates of the ENE were not filtered for those who had worked at least one month in their current job, and the expansion factors of each survey were different. Because the period of analysis for the ESI is longer than that of the ENE (the ESI is published annually, the ENE monthly), a more in-depth review of the income module might find classification errors in the employment module.

In 2013, the criterion of keeping all information in the employment module unchanged was therefore relaxed and, through an extensive analysis of the occupational profile, changes may be made in the ENE module. The relaxation of the criterion has provided a new perspective on the differences between ENE and ESI estimates of the number of employed persons.

For example, the 2013 ESI corrected a misclassification of employed persons who had been classified as own-account workers when in fact they should have been classified as employers. Even more important than the totals, this change opens a possibility, which did not exist until the 2012 version, of improving the precision in the measurement of personal income.

- **Use of a reference table and its effect on transfers related to education**

Experience in conducting income surveys shows that respondents often do not know the amount of their benefits or underestimate their value. This problem is common when we include in-kind transfers such as food grants or scholarships. The international recommendation is to use a reference table in the field to circumvent underestimation.

In field evaluations of the first version of the ESI (2010), the use of the reference table by interviewers was slow because they sometimes forgot that they had this tool at their disposal and left the estimate of the transfer's value to the respondent. This situation resulted in a retraining plan for survey teams and the preparation of instructions on the use of the reference table, which resulted in greater understanding and use of the instrument in the fieldwork of the 2011 ESI.

Learning how to use the reference table resulted in a considerable increase in the number of "scholarships" captured in 2011 compared to 2010 and in an increase of 76.4% in the number of people who reported the benefit in 2011 compared to 2010.

Therefore, if the item "scholarships" is included in household income, the income structure of households in 2011 *is not comparable* to that of 2010. The measure taken was to publish tabulations with and without "scholarships" in the household transfers component in order to have tabulations in 2011 that were comparable with those of 2010 by isolating the effect of the increased capture of "scholarships".

Because the results from the 2012 ESI to the 2020 ESI are comparable with those of 2011, the analyses can be made with or without educational subsidies, according to objectives, and both tabulations are published on the institutional website. Nevertheless, the documents with the main results officially produced through the ESI refer to the version that does not include scholarships.

- **Change in the classification of branches of economic activity**

Based on the need to include the new productive and technological capabilities and to mitigate the classification problems that currently arise in sociodemographic surveys (household surveys), INE has moved from the International Standard Industrial Classification of All Economic Activities (ISIC) rev.3, to ISIC rev.4, the latter adapted to Chile (CIIU rev4.cl CAENES).

To smooth the transition, the tabulations from 2013 to 2016 were published with both classifications. Since 2017, the official classification of economic activities has been CIIU rev4.cl CAENES.

2.1.4   Information needs

The purpose of the ESI is to respond to the needs of academics, researchers, students, and the general public for timely information on labour income of the country and its regions year after year. The information provided by the survey can also assist the State in making decisions that benefit the entire population.

2.2 Objectives of the study

General objectives

- To characterize the labour income in the reference month of persons who are classified as employed in the ENE.

- To characterize income from jobs other than the main job for persons who are classified in the ENE as employed and not employed.

Specific objectives

- To characterize the income from the main job, according to year, sex, employment classification, educational level, employment status (formal and informal), occupational group, and branch of economic activity, among other characteristics, at the national and regional levels.

- To characterize the income from the main job of own-account workers, according to sex, workday, place of work, and social security contributions, among other characteristics, at the national and regional levels.

- To characterize the income from the main job of paid employees, according to sex, enterprise size, minimum net income bracket, and hourly income, among other characteristics, at the national and regional levels.

- To characterize income from other jobs.

III. Reference framework

3.1 Theoretical-conceptual framework

The theoretical-conceptual basis of the survey has two axes: one related to its social relevance and the other to its conceptual and methodological guidelines.

3.1.1 The relevance of and international guidelines for the compilation of income statistics

The ESI responds to the need for a comprehensive national description of labour income and income from occupations other than the main job, and it complements the employment statistics of the ENE through its characterization of the income of the nation's employed population.

The development of the ESI involves a set of international recommendations and discussions on the desirability and social relevance of compiling and standardizing information on income:

- International guidelines, such as those found in the Canberra Group Manual (United Nations, 2011), mention that the compilation of income-distribution data is relevant to the extent that it illuminates income distribution patterns and their relationship to the organization of societies, and it thus enables the evaluation of the impact of public policies in areas of interest and the analysis of how different income distribution patterns influence household welfare and people's ability to acquire the goods and services that satisfy their needs.

  Furthermore, the manual mentions that the focus on income distribution may be justified based on its observation of how the benefits of the national product are distributed among the population or based on its use as an indirect measurement of the distribution of economic well-being (United Nations, 2011).

- For its part, the International Labour Organization (ILO) provides guidelines through the "International Conference of Labour Statisticians" (ICLS). The resolution of the 16th ICLS (OIT, 1998) mentions the importance for countries of complementing their statistical programs on employment, unemployment, underemployment, and wages with information on employment-related income in order to analyze the income-generating capacity of economic activities and to analyze the economic well-being of persons based on employment opportunities.

- The 17th ICLS emphasizes that household income and expenditure statistics can serve as a basis for the descriptions of various economic, social, and other characteristics of households, among which are an assessment of the level, structure, and trends in the economic well-being of households and persons that illustrates how income contributes to the potential of persons to consume or save (OIT, 2003).

3.1.2 Conceptual and methodological guidelines for the compilation of income statistics

Income

Specifically, the ESI measures sources of income that fall under the concept of **current income**: income from the main job and income from occupations other than the main job during the reference period.

- Income from employment (from jobs): net receipts from paid and self-employment jobs.

- The definition of current income is conceptually framed within the system of national accounts (SNA), where it is mentioned that, from a theoretical point of view, income is considered as the "maximum amount that a household or other unit can afford to spend on consumption goods or services during the accounting period without having to finance its expenditures by reducing its cash, by disposing of other financial or non-financial assets, or by increasing its liabilities." (GITCN, 2009). This definition would exclude income such as capital transfers, real holding gains or losses, and other changes in the volume of assets resulting from such events as natural disasters.

Although the approach responds to different analytical interests (macroeconomic in the case of the SNA and microeconomic in the case of statistics on personal-income distribution (United Nations, 2011)), common guidelines that mark the concept of income can be established:

- It refers to regular or periodic income rather than extraordinary income.
- This measurement is oriented towards current expenditure and, therefore, **net income** from employment (paid and self-employment jobs).
- It thus excludes any transactions involving changes in the units' assets.

Components of income measurement

As stated in the 17th ICLS (OIT, 2003) and by the Canberra Group (United Nations, 2011), income from

- Jobs (paid or self-employment)

(a) Income from employment or occupation

Underlying employment and income is the **concept of work**. Although in practice both are used as synonyms, the meaning of work is broader than that of employment. Work refers to "the set of human activities, paid or unpaid, that produce goods or services in an economy, or that satisfy the needs of a community or provide the necessary means of support for individuals" (OIT, 2004). On the other hand, the **concept of employment** refers to work performed in exchange for payment, regardless of the dependency relationship between the parties (OIT, 2004).

In this sense, the concept of **employment** is of special relevance for the measurement of personal income, and international recommendations have thus been established for adequate recording.

In the resolution of the 16th ICLS, as also noted by the Canberra Group, income from employment consists of income from participation in economic activities in a strictly job-related context (OIT, 1998). Income from employment consists of payments in cash or in kind received by individuals for themselves or by their family members, as a result of their current or future involvement in paid or self-employment work (United Nations, 2011).

The reference framework for income measurement is limited, at a general level, by the so-called **SNA production boundary**, which in general involves the activities (physical processes) carried out by an institutional unit (including households, persons, or groups of persons) to transform inputs of goods and services into outputs of other goods and services that are destined for the market through sale or barter. This conception of the production boundary considers the possibility of non-monetary transactions between different units of the economy and even within the same unit, such as a household (with the exception of services for own consumption, as will be mentioned in a later paragraph).

According to the 19th ICLS resolution, a job or work activity "is defined as a set of tasks and duties performed, or meant to be performed, by one person for a single economic unit" (OIT, 2013). This implies that a person may have one or more jobs. Among these jobs, the **main job** may be defined as the one in which the person usually works the most hours.

As described by the Canberra Group and agreed upon in the ICLS, income from employment includes the following:

- **Income from paid employment**, received in cash or in kind (goods or services), consists of direct wages, salaries, and other benefits provided by the employer, which should be recorded separately when appropriate. Payments for goods or services are valued and included only to the extent that they are unconditional, otherwise they are excluded or valued at zero.

- **Income from self-employment**, which is linked to income received as a result of engaging in self-employment and is therefore directly dependent on the benefits derived from the goods and services produced. Self-employed workers are so defined because they are, unlike paid employees, directly dependent on the welfare of the enterprise (OIT, 1993, p. 2). In accordance with the SNA definition, income from self-employment is mixed income because it is not possible to distinguish the

self-assignment of wages from the profits obtained by the owner as an entrepreneur (GITCN, 2009).

3.2 National and international benchmarks

Table 2 below summarizes the main characteristics of statistical products similar to the ESI in terms of subject matter (personal income) and methodology (face-to-face household surveys of a probabilistic nature), which have been conducted in Chile and other countries.

The criterion for the inclusion of other countries was a similar per capita income within South America[11], to which Mexico was added because it is the only other Latin American member of the OECD besides Chile.

---

[11] Accessed at https://estadisticas.cepal.org/cepalstat/

Table 2 National and international benchmarks related to the ESI

Country	Survey and Frequency	Key indicators; Survey and Respondent Population	Sampling design and Sample size	Collection period and reference period	Instrument	Observations
Argentina	Permanent Household Survey (EPH) - INDEC[12]	- Household characteristics, demographic characteristics of persons - Employment characteristics and migration - Housing, education, and income characteristics - Households and persons - Individual questionnaire provides information on persons aged 10 and over	Two-stage stratified samples, primary units (census radii) grouped in strata, selection proportional to size (number of dwellings). Second stage: dwellings 52,340 surveys first quarter 2018	Results are delivered four times a year: quarters Reference period is weekly, except for monthly indicators	Three questionnaires: - Questionnaire for individuals: **employment and income characteristics** - Questionnaire for dwellings - Questionnaire for households	Questionnaire for individuals: - A set of questions designed to inquire about the different sources of income, both for persons and households as a whole. - The income surveyed comes from both labour sources (paid employees and self-employment workers) and non-labour sources (mainly pensions and subsidies). - Includes questions on interurban movements in employment. Questionnaire for households: - Includes questions on non-labour income. - Contains a module of questions on housework. - Module on "household maintenance strategies", income earners
Brazil	Continuous Monthly Household Survey (PNAD) - IBGE[13]	- Employment, integration of people into the labour market - Persons - The individual questionnaire provides information on persons aged 10 and over. - Persons of working age are interviewed: aged 14 and over.	Probability sample 211,344 permanent home addresses per quarter	Results are delivered four times a year: quarters Reference week for employment For income, reference month prior to the week	Single questionnaire	- The questionnaire has a module for the characterization of persons, as well as a module for the characterization of their main, secondary, and other jobs. - The logic of the questionnaire integrates questions by main, secondary, and other jobs with income according to type, as well as other complementary income provided by the employer or for own consumption. The questionnaire then asks about hours, under-employment, and unemployment.

---

[12] Accessed at https://www.indec.gob.ar/uploads/informesdeprensa/eph_pobreza_01_18.pdf.

[13] Accessed at: https://www.ibge.gov.br/estatisticas-novoportal/sociais/trabalho/9171-pesquisa-nacional-por-amostra-de-domicilios-continua-mensal.html.

					- Includes own-consumption activities (services). Includes a module on volunteer work and caregiving. Includes a module on housekeeping. - Consultation on gross income from work - Consultation on discounts for delays, absences, etc. Consultation on how discounts from income are calculated. - Includes basic modules on property income. - Includes more detailed module on transfers. - Includes consultation on imputed rent.
Chile	CASEN Survey – Ministry of Social Development[14]	Demographics, education, health, housing, labour, and income Persons and households Head of each household, otherwise a household member aged over 18.	Two-stage probability sample in urban and rural areas. Classification of urban blocks according to size (size strata). Selection of dwellings with equal probability in block (urban) or section (rural) strata. 2017 sample: 70,948 households, 216,439 persons	Provides results every two years Reference to employment in the previous week, month Reference to income in the previous month	Single questionnaire	- Questionnaire includes questions on income from work, property, and transfers. - Employment includes main activity, other jobs. - In addition to consultation by reference month, the questionnaire includes a measurement of income over twelve-month periods.
Mexico	National Survey of Household Income and Expenditure – ENIGH[15]	- Income - Expenditure - Economic activities carried out by household members - Target population: households - Suitable respondent: head of household, homemaker, or any member of the household aged 18 or	Probabilistic, stratified, and clustered (two-stage) design. The ultimate unit of selection is the dwelling and the unit of observation is the household. 81,515 dwellings nationwide, with a 90%	The survey is conducted every two years. In the specified year, the survey is conducted over three months, which are divided into nine periods of ten days.	Various questionnaires: - Questionnaire for households and dwellings - Daily household expense booklet - Questionnaire on household expenditures - Questionnaire for persons aged 12 and over: employment, income from main job, income from secondary job, income	- Questionnaire for persons aged 12 and over includes measurement of items related to employment, as well as income (**current income**), both from work and from other sources. - The previous questionnaire also includes the measurement of income that does not meet the definition of current income: financial and capital receipts that modify the net value of the household's assets and do not meet the criteria of regularity and destination.

---

[14] Accessed at http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/docs/Diseno_Muestral_Casen_2017_MDS.pdf.

[15] Accessed at https://www.inegi.org.mx/programas/enigh/nc/2016/default.html

		over who knows the household information as a whole. - Ideally, all persons aged 12 and over provide their own information. Persons aged under 12: a report of a person aged 18 or over who knows the information.	confidence level and a maximum relative error of 7.20%.	**b. Various** For household income module for persons aged over 12: month prior to the month of interview. It also captures income from up to six months prior to the reference month. For children aged under 12, the survey includes income up to six months prior to the reference month.	from other jobs in the past month, non-labour income, social networks, time use, health. - Household business questionnaire: household business characteristics and income, consistency of information - Questionnaire for persons aged under 12: income, access to healthcare.	- Includes questionnaire that collects information from household businesses, when applicable. - In addition, there is a questionnaire for persons aged under 12, and it records their income based on the respondent head of household or person in charge who is aged 18 or over and who knows the information. - Includes questionnaires and booklets for expenditures. - Income from work includes the measurement of workers' compensation, severance payments, and voluntary retirement. - The survey excludes from self-employment income any income received in salaries or profits of workers who are partners in an incorporated enterprise and of workers who are owners of enterprises that, although not incorporated, carry out full accounting practices (quasi-corporations). - It also considers the measurement of "mixed income" in self-employment. - Expressly considers the measurement of royalties (copyrights, patents) in property income.
Mexico	National Survey of Occupation and Employment (ENOE) - ENIGH[16]	- Occupation and employment - Income - Data are published for the population aged 15 and over and for households. Economic	The type of sampling used is probabilistic, two-stage, stratified, and clustered. Quarterly: 120,260 surveys.	Quarterly. Monthly and quarterly measurements. Sociodemographic characteristics: time of interview.	- Sociodemographic questionnaire. - Jobs and Employment Questionnaire, basic and extended	The survey is based on the regulatory framework of national accounting systems. - It notes in the definition of economic activity that, if a service is not requested, it is not a real transaction and therefore is not an economic activity, because there is no demand for it in society. The survey draws

[16] Accessed at https://www.inegi.org.mx/programas/enoe/15ymas/.

		data is collected on persons from ages 12 to 14. - Suitable respondent: head of household or other household member aged 15 or over. Ideally, all persons aged 12 and over provide their own information.		Economic characteristics: previous week (Monday to Sunday).		this distinction, and it is not left to the discretion of the respondent whether an activity is defined as an occupation. Example: windshield cleaners and those who beg for alms. - "Employed persons" are defined as such under the premise that someone is demanding what they offer. - Capture of income from main job. The survey does not capture income from secondary or other occupations, nor does it capture income from property or from transfers.
Uruguay	Continuous Survey of Households[17] - INE	- Official labour market indicators (activity, employment, and unemployment) - Personal and household income with - Poverty and indigence thresholds on an annual basis. - Basis for social research on other topics. - Households and all persons in households in selected dwellings. Employment and income: persons aged 14 and over. - Respondent is a household member aged 18 or over (except domestic service) who is mentally competent to provide information	Complex sampling design that includes several selection stages: stratification of departments and localities according to size; selection of census areas; selection of dwellings. 53,580 cases per year, 4,465 cases per month.	Monthly, quarterly, semiannual, and annual frequency. The reference period is the week prior to conducting the interview: employment.	Single questionnaire: - Dwelling and household characterization module - Personal data	- Collected with electronic devices. - The research on the labour force and earned income applies only to persons aged 14 and over. This threshold is the same as the minimum age of the Economically Active Population (EAP). - Measures current household income from work, property, transfers, etc. - Expressly considers the measurement of royalties (copyrights, patents) in property income. - It specially distinguishes the income of agricultural workers. It considers possible income from "sharecropping", "herding", "livestock to capitalization". - Self-employed workers are asked a specific question to specify received income as a share of profits in order to contrast with another question that asks if money was withdrawn from own businesses for household expenses.

[17] Accessed at http://www.ine.gub.uy/encuesta-continua-de-hogares1.

		about the other household members.			
Uruguay	Survey of Household Income and Expenditure - ENGIH 2016/2017[18] - INE	- Information on households and dwellings - Characteristics of persons: education and **work**, **income**, household and household members' expenditures - Analysis at the level of single household, dwelling, household members. - Each member of the household is interviewed in person.	Random sampling Stratified with optimal allocation. Random, stratified, and clustered, in two or three stages. Sub-universes, geographical strata/ area, dwellings with private households, households 7,500 dwellings, all household members	The survey is conducted every ten years. The latest version was conducted throughout 2017. Weekly and monthly reference periods (monthly income)	Four instruments: Questionnaire 1: - Information on dwelling and household, and sociodemographic and other social information on persons. - Personal and household income. - Household expenditure. - Daily household expenditure. Questionnaire 2: daily household expenditure Questionnaire 3: personal daily expenditures Questionnaire 4: household expenditure	Questionnaire 1: - Measures current household income from work, property, transfers. - Expressly considers the measurement of royalties (copyrights, patents) in property income. - It specially distinguishes the income of agricultural workers. It considers possible income from "sharecropping", "herding", "livestock to capitalization". - Self-employed workers are asked a specific question to specify received income as a share of profits in order to contrast with another question that asks if money was withdrawn from own businesses for household expenses.

[18] Accessed at http://www.ine.gub.uy/engih2016.

IV.   Statistical design[19]

4.1 Universe

Every year, the ESI collects information on the income of persons who reside in occupied private dwellings, who are present in the national territory, and who are covered by the sampling frame defined for the survey (the 2017 Master Sampling Frame of Dwellings). The ESI also gathers data on income associated with a main job[20], which must be the same as the main job of the previous month.

4.2 Target population

The target population is the working-age population (i.e., all persons aged 15 and over who habitually reside in occupied private households in the nation).

4.3 Statistical units

- **Unit of analysis**

In order to meet the objective of the survey, the units of information and analysis are defined as persons who form part of the target population, that is, those aged 15 or over who habitually reside in households located in private occupied dwellings within the national territory. However, the sociodemographic characteristics of each resident of the selected dwelling are recorded through a questionnaire designed for this purpose.

- **Sampling unit**

There are two sampling units.

(a) The primary sampling unit (PSU) is defined as a homogeneous cluster in terms of the number of private dwellings of which it consists.

(b) The secondary sampling unit (SSU) is defined as a private occupied dwelling within the selected PSUs.

---

[19] For more on the statistical design, see https://www.ine.gob.cl/docs/default-source/employment-and-unemployment/methodologies/documents/methodology-ene-2020.pdf?sfvrsn=c3e5bc29_10.

[20] This data characterizes the labour income of persons who are classified as employed in the ENE.

The average size of PSUs in the urban area is 200 private dwellings, with a range from 160 to 240, excluding seasonal dwellings. The average size of rural PSUs is 90 private dwellings, with a range of 70 to 110, excluding seasonal dwellings.

4.4 Variables measured

Current personal income from occupation.

The variable measured in the ESI is the following:

- **Labour income (income from occupation)**

  - Income of dependent employed persons (including paid employees and domestic workers)
  - Income of self-employed persons (including own-account workers and employers)
  - Income from other work

  For income from employment, the following points should be considered:

  1.1 The measurement of the **employed population** uses demographic information from the month prior to the reference week of the ENE.

  According to the ESI, consequently, the employed population refers to the group of people for whom their **current job is the same job they had in the reference month** and, therefore, all profiles that characterize employment are extended to the analysis of the labour income of this group of people.

  With the survey, we can identify the population of those who had a second job in the previous month or a job different from the one they had in the reference week of the survey (including those who were not necessarily employed in the reference week but who worked in the previous month) and who reported information on labour income. For this group, employment profiles that may have been reported in the core questionnaire of the ENE cannot be extended to the ESI[21], because the only job included in the ENE is the main job performed in the reference week.

  In summary, the information on employment income of employees and self-employed persons refers exclusively to those whose current job is the same as the one they had in the reference month.

---

[21] Income from secondary occupations is captured in a specific item of the ESI module. It is a comprehensive question that encompasses cash and in-kind income.

1.2 The ESI measures income of self-employed persons according to two occupational categories:

The first is "employer", which is defined as a person who is owner of any enterprise or business and who hires one or more persons as paid employees (unpaid family members are excluded in this case) over a continuous period that includes the reference period.

The second is "own-account workers", who are persons working on their own account. This category constitutes the second largest group within the employed population but, unlike employers, they do not continuously hire other persons as paid employees to work with them.

For the self-employed, whether or not they are employers, the objective is to separate gross income from total expenses in order to obtain, as in the case of paid employees, the net income received for the completed activity. Because self-employed persons receive income from capital and work, what is reported in the survey is a combination of profits and salaries[22], which is referred to in the literature as "mixed" income[23].

4.5 Nomenclature used[24]

Classification of Status in Employment

Classification of the employed includes the use of the International Classification of Status in Employment (ICSE), which was adopted by the International Labour Organization at the 15th ICLS in 1993 (OIT, 1993). This system classifies the employed population into categories such as employer, self-employed, paid employee, unpaid family worker, and domestic worker.

National Classification of Economic Activities

The National Classification of Economic Activities for Sociodemographic Surveys (CAENES), which was developed by the National Statistics Institute of Chile, is the result of the adaptation

---

[22] For example, when owners of an enterprise or business assign a salary to themselves and receive a share of the enterprise's profits.

[23] The term "mixed" income is a way of indicating the difficulties of distinguishing capital depreciation and profits from the payment of wages and salaries.

[24] The information contained in this section is based on what is stated in the metadata referring to the National Employment Survey, which were retrieved from the web page (https://www.ine.gob.cl/estadisticas/sociales/mercado-laboral).

of the Chilean Standard Industrial Classification CIIU4.CL 2012. This classification tool includes criteria that resulted from an exhaustive analysis of the most frequent inquiries during the coding process of the current ENE (INE, 2016).

The CAENES structure has three hierarchical levels: *section*, which is identified by a letter (identical and equivalent to the sections of CIIU4.CL 2012); *division*, which consists of two numerical digits, corresponding in most cases to the same categories of CIIU4.CL 2012; and *class*, which consists of four digits. For classification, the representativeness of the estimates provided by the ESI covers all twenty-one sections of the CIIU (ISIC, by its acronym in English).

International Classification of Occupations

The classification of occupations is based on the International Standard Classification of Occupations (ISCO-88), which was published by the International Labour Organization in 1988. ISCO is "a tool for organizing jobs into a clearly defined set of groups according to the tasks and duties undertaken in the job"[25].

The ISCO-88 consists of ten major groups, twenty-eight sub-major groups, 116 minor groups, and 360 unit groups. The first digit indicates the major group, and the second, third, and fourth digits represent the sub-major group, the minor group, and the unit group, respectively. For classification, the representativeness of the estimates provided by the ESI is at the level of major groups.

The use of the Chilean Classification of Occupations, CIUO-08.CL, the Chilean adaptation of the International Standard Classification of Occupations (ISCO-08), began with the 2018 ESI.

Classification of Education

The 1997 International Standard Classification of Education (ISCED), conceived by UNESCO in 1970 and subsequently revised, is used to classify the population according to educational level. Its update was approved by the UNESCO General Conference at its 29th session in November 1997. The design of ISCED 1997 included consultations with representatives from around the world.

The purpose of ISCED is to collect, compile, and present comparable indicators and statistics on education both within a country and internationally. This classification covers two variables: levels and fields of education (UNESCO, 1997).

---

[25] Information accessed at http://www.ilo.org/public/spanish/bureau/stat/isco/index.htm.

4.6 Design of the instrument for information capture

4.6.1 Questionnaire design

The survey questionnaire[26] has a unique design that is applicable to any reporting unit. It consists of six modules:

- Module A Employment status in the reference week
- Module B Main activity
- Module C Hours worked
- Module E Search for employment
- Module D Income
- Module G Characteristics of the household

The unit of information (the eligible respondent) is defined as a person who is a member of the surveyed household, who usually resides in the dwelling, who is aged 15 or over, and who is in a position to answer the questionnaire with information on the respondent, or on a part or all of the members of the household.

4.6.2 Pilot test

In August 2010, a pilot survey was conducted as part of the New National Employment Survey (NENE). When revising the previous Employment Survey, the income module, which is applied in the last quarter of the current year, was updated for the NENE. Within this framework, the pilot test sought to field-test the questionnaire developed by the technical team in charge of the NESI (New ESI), which based its underlying theory on the following:

- The provisions and suggestions of the **Canberra Group** (The Canberra Group, 2001).
- **The Sixteenth International Conference of Labour Statisticians,** Report on Labour Income (OIT, 1998).
- The UN **System of National Accounts** (GITCN, 1993).

The purpose of the pilot test was to enable the new questionnaire to capture mixed income, to reinforce the macroeconomic approach to household income, and to safeguard the microeconomic approach to the analysis of income distribution and well-being of households and persons.

In operational terms, the survey was conducted over two weeks in August 2010 in the Metropolitan Region. The ESI questionnaire was attached to the central NENE form, and the

---

[26] The form is available (in Spanish only) on the website www.ine.cl, in the section *Ingresos y Gastos de las personas* (Personal Income and Expenditure) at https://www.ine.gob.cl/estadisticas/sociales/ingresos-y-gastos/encuesta-suplementaria-de-ingresos, under the heading "*formularios*" (forms).

questionnaire was administered immediately after the NENE questionnaire (sections A, B, C, E).

With the results obtained from the pilot survey, the following general recommendations were established for the application of the instrument:

- The NESI was well received, and a preference was noted for including it as a module within the NENE instead of its application during a second visit.
- The NESI enabled information on high-income households to be collected.
- The results of the NESI showed that there were no major disparities in the imputed rent estimates for households on the same block or street.
- An integrated system of income surveys was proposed, a system that emphasizes Household Budget Surveys (instead of short-term surveys) in order to avoid biases in the delivery of information.

More specifically, improvements were made to the income questionnaire that determined its incomparability with versions of the ESI prior to 2010. The improvements, however, enabled greater precision in the measurement of labour and household income, as also described in this document.

4.7 Sampling methodology[27]

An explanation of the sampling design in English can be found at www.ine.cl, section Labour Market, sub-section Employment and Unemployment, Methodologies, "Methodology ENE 2020"[28].

As mentioned above, the ESI shares all the attributes of the ENE sampling design. It therefore also shares the definition of the theoretical sample size for the months of October, November, and December, the quarter in which the ESI is conducted.

As a result of the COVID-19 pandemic in Chile, the National Statistics Institute adopted a series of measures to give continuity to the collection of ENE data and to protect its interviewers and respondents from unnecessary exposure. These measures were adopted in March 2020 and will continue until they are no longer necessary.

---

[27] The information contained in this section is based on metadata on the National Employment Survey, which can found at the following web page: https://www.ine.gob.cl/estadisticas/sociales/mercado-laboral/ocupacion-y-desocupacion, under the heading "*metadatos*".

[28] An English translation of this methodological document can be found at https://www.ine.gob.cl/docs/default-source/employment-and-unemployment/methodologies/documents/methodology-ene-2020.pdf?sfvrsn=c3e5bc29_10.

Attainment rate for the target sample[29]

Table 3 shows the attainment rates with respect to the target sample at the national level during the October-November-December 2021 (OND 2021) moving quarter. The attainment rate for the December 2021 subsample was 86.1%, which is equivalent to 11,387 dwellings nationwide. For the moving quarter, the attainment rate was 88.3%.

Table 3 Attainment rate for the ENE compared to the target sample

Subsample	Percentage	No. of dwellings attained
Oct 2021	86.8%	11,479
Nov 2021	91.9%	12,143
Dec 2021	86.1%	11,387
Moving quarter October-November-December 2021 (OND 2021)	88.3%	35,009

Source: INE

4.7.1   Measurement of precision

Sampling errors result from using a part of the population to estimate characteristics of the whole population. However, these errors for the variables of interest can be measured by different methods, such as variance, standard deviation, or other forms of these methods, such as absolute error, coefficient of variation, and relative error. The ESI primarily uses the last three methods.

Absolute Error and Confidence Interval

The absolute error is used to calculate the confidence interval of the estimator with 95% confidence, $\hat{Y}$, and it is the estimation error expressed in the same units as the variable Y under analysis[30].

The absolute error associated with an estimate is defined as the maximum expected difference between the population value and the estimated value, given a confidence level of 95%, and it determines the confidence interval of the variable of interest Y. The absolute error is obtained as the product of the standard deviation $(S(\hat{y}))$ of the estimate and the critical value associated with the percentile of the t-distribution with sixty degrees of freedom

---

[29] Because we are in the process of incorporating the new sampling design, the attainment rate is calculated with respect to the target sample for each subsample. The number of responses obtained from dwellings are defined as responses that pass a sufficiency matrix (quality control) and enter into the final calculation.

[30] When the estimator $\hat{Y}$ is a percentage or rate, the absolute error is expressed in percentage points.

$(t_{1-\frac{\alpha}{2},\ 60})$, where the first concept refers to the confidence level for the desired inference (95%) and the second concept refers to the sixty degrees of freedom with which the critical value is calculated[31].

This can be expressed in the following manner:

$$E_a(\hat{y}) = t_{1-\frac{\alpha}{2},60} \cdot S(\hat{y})$$

In probabilistic terms, it can be expressed as follows:

$$\Pr\left(\left|\hat{Y} - Y\right| \le E_\alpha(\hat{y})\right) = 1 - \alpha$$

Where $\hat{Y}$ and Y are the estimated and true values of the population, respectively.

The 95% confidence interval for the true value of the variable Y in the population is calculated as follows:

$$CI(\hat{y})_{95\%} = [\hat{y} - E_\alpha(\hat{y}); \hat{y} + E_\alpha(\hat{y})]$$

$$CI(\hat{y})_{95\%} = [\hat{y} - t_{1-\frac{\alpha}{2},60} \cdot S(\hat{y}); \hat{y} + t_{1-\frac{\alpha}{2},60} \cdot S(\hat{y})]$$

- **Coefficient of Variation and Relative Error[32]**

The relative error is expressed as a percentage, and it refers to the maximum difference between the point estimate and the true value of variable $\hat{y}$ in the population, with 95% confidence, as indicated by the t-distribution percentile associated with sixty degrees of freedom[33] $\left(t_{1-\frac{\alpha}{2},60}\right)$.

Its mathematical expression for a variable $(\hat{y})$ is as follows:

---

[31] For the 2020 ESI and 2021 ESI, the degrees of freedom were set at sixty for the calculation of the absolute error and the confidence interval in all subpopulation estimators. The decision to use this level in the degrees of freedom was based on having a dual sampling frame, which made it difficult to locate the observations in strata and single clusters. This condition does not apply to the evaluation of the estimate quality, where the degrees of freedom are obtained in the traditional way. (See the next point on the evaluation of the quality of estimates.)

[32] For more details on the calculation method using a variety of statistical software, see Annex 2.

[33] See note 34.

$$E_r(\hat{y}) = t_{1-\frac{\alpha}{2},\ 60} \cdot \frac{\sqrt{S(\hat{y})}}{\hat{y}}$$

Where $\frac{\sqrt{S(\hat{y})}}{\hat{y}}$ represents the coefficient of variation associated with the estimation of variable $\hat{y}$.

Both the relative error and the coefficient of variation can be expressed as percentages. For example, for a relative error of 20% at a 95% confidence level with sixty degrees of freedom (critical t-value = 2.0002978), the true population value of the variable is plus or minus 20% of the estimated value, with a confidence level of 95%.

The interpretation of the coefficient of variation is identical: the relative error is a monotonic transformation of the coefficient of variation because it is multiplied by the scalar $\left(t_{1-\frac{\alpha}{2},60}\right)$.

Thus, for a confidence level of 95%, the coefficient of variation is approximately half the relative error in sufficiently large samples.

- **Hypothesis testing[34]:**

Hypothesis testing allows us to confirm, with a certain level of confidence, whether the difference between two estimates is statistically significant. Since the 2020 ESI, these statistical tests[35] have been calculated as the product of the standard deviation of the difference estimate (obtained through the delta method) and the critical value associated with the percentile of the t-distribution with sixty degrees of freedom (see footnote 33). This section delves into the statistical concepts associated with these tests.

A hypothesis test, in broad terms, is a procedure for testing two hypotheses or statements about the value of a certain parameter θ in the population or the difference of two population parameters $\theta_1 - \theta_2$ from two subpopulations, respectively.

All hypothesis testing is based on the formulation of two exhaustive and mutually exclusive hypotheses (the null hypothesis and alternative hypothesis).

- The null hypothesis (H_0) assigns a value or range of values to the parameter in question. It is the statement that can be rejected by facts or empirical evidence through a sampling procedure.

---

[34] See Annex 5 for hypothesis testing that compares different years of the ESI.

[35] These statistical tests refer specifically to measurements of the differences in means, totals, proportions, and ratios within a single year. Tests of median difference have been excluded because their measures of variability are obtained from repeated sub-bootstrap designs whose methodological implications are being evaluated for incorporation.

- The alternative hypothesis (H_1) is the negation of the null hypothesis and the alternative of interest on which the researcher focuses.

Statement of contrasting hypotheses:

When comparing totals, proportions, means, and ratios (without losing generality), the parameter θ can be used as a reference as if it were the population mean, although the difference of any of the other descriptive statistics mentioned above can be contrasted.

$H_0: \theta_1 - \theta_2 = 0 \iff H_0: \theta_1 = \theta_2$	No difference between population means
$H_1: \theta_1 - \theta_2 \neq 0 \iff H_0: \theta_1 \neq \theta_2$	Difference exists between population means

The test statistic

The test statistic is a sample result that meets the following conditions:

- It provides relevant empirical information on the statement proposed in the null hypothesis (H_0). Estimated differences (across samples) relatively distant from zero provide evidence against H_0 and in favor of H_1.

- It has a known sampling distribution. For the null hypothesis (under the assumption that population means are equal) and for samples large enough to apply the central limit theorem, the sampling difference is approximately normally distributed if the population variance of the difference is known.

- When this population variance is unknown (in almost every case), and it is estimated across the same samples, the standardized difference follows a t-distribution, with v degrees of freedom[36].

For two-stage cluster and stratified sampling, the v degrees of freedom of the t-distribution correspond to the number of non-repeated clusters sampled in the evaluated subpopulation minus the number of non-repeated strata sampled in the evaluated subpopulation. As of 2020, there are two sampling frames (dual frames), whose old and new clusters do not have the same geographical configuration nor do they have the same strata (inputs with which the degrees of freedom are calculated). Although they do not have the same geographical configuration, there is partial and complete overlapping of the clusters and strata, so the real degrees of freedom of the contrasts are being studied. Applying a semi-conservative criterion, we decided to use $v = 60$ degrees of freedom. With this decision, the percentile of the t-distribution corresponding to the bilateral test with a significance level $\alpha = 5\%$, is $t^v_{1-\alpha/2} = t^{60}_{0,975} = 2.000297822 \cdots \approx 2$.

---

[36] For hypothesis tests involving any estimator of the 2020 or 2021 ESI, the number of degrees of freedom is sixty (see footnote 33).

The standardized test statistic is calculated as follows:

$$t = \frac{(\hat{\theta}_1 - \hat{\theta}_2) - (\theta_1 - \theta_2)_0}{\sqrt{\hat{V}(\hat{\theta}_1 - \hat{\theta}_2)}} = \frac{(\hat{\theta}_1 - \hat{\theta}_2)}{\sqrt{\hat{V}(\hat{\theta}_1) + \hat{V}(\hat{\theta}_2) - 2 \cdot \overline{COV}(\hat{\theta}_1 ; \hat{\theta}_2)}} \sim t^{60}, \text{ where } H_0, \text{ that is, } (\theta_1 - \theta_2)_0 = 0$$

Note that the estimated variance of the difference includes the estimated covariance between the means of the underlying populations because, in complex cluster designs, variances are calculated by the ultimate cluster method and samples may share common clusters, strata, or domains. Statistical software such as Stata, R[37], SAS, and SPSS employ the Delta method or Taylor linearization that includes the covariance term.

Critical region H_0 (CR):

For a difference hypothesis test, a critical region of the null hypothesis is established, so that if the observed difference $(\hat{\theta}_1 - \hat{\theta}_2)_{obs}$ falls within this CR, the null hypothesis of equality is rejected, thus establishing statistically that there is a difference and that the population difference of the statistics is significant. In figure 1, the observed difference does not fall within the CR. Instead, it falls in the region of acceptance of the null hypothesis of size $1 - \alpha$. Thus, it can be concluded that there is no evidence to reject H_0, that the difference is not significant, that there is no difference between the population statistics, and that the population statistics are equal.

Figure 1 Critical region of level α for the bilateral hypothesis test
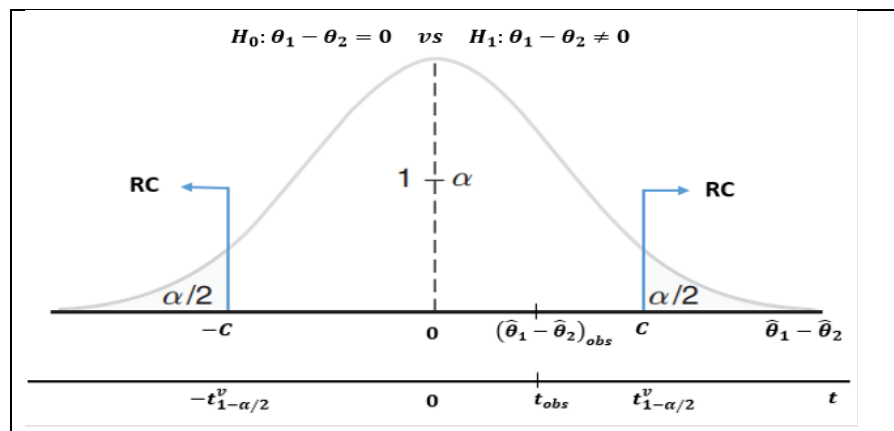
Source: INE

The CR is formed by the set of all absolute differences observed on the axis of the graph $\hat{\theta}_1 - \hat{\theta}_2$ that are greater than the critical value $C > 0$, or equivalently, formed by all standardized absolute differences (values on the t axis of the figure) that are greater than the percentile of the t-distribution with v degrees of freedom, which, given the symmetry of this distribution, is the percentile of the order $1 - \alpha/2$, leaving under the curve an area with size $\alpha/2$ for each end of the distribution.

---

[37] In the case of R, these techniques used the *svycontrast* function of the *survey* package.

$$CR = \left\{ \hat{\theta}_1 - \hat{\theta}_2 \; ; \; \left| \hat{\theta}_1 - \hat{\theta}_2 \right| > C \right\} \Leftrightarrow \left\{ \hat{\theta}_1 - \hat{\theta}_2 \; ; \; \hat{\theta}_1 - \hat{\theta}_2 < -C \; \text{ó} \; \hat{\theta}_1 - \hat{\theta}_2 > C \right\} \Rightarrow Prob(CR) = \alpha$$

$$CR = \left\{ t \; ; \; |t| > t^v_{1-\alpha/2} \right\} \Leftrightarrow \left\{ t \; ; \; t < -t^v_{1-\alpha/2} \; \text{ó} \; t > t^v_{1-\alpha/2} \right\} \Rightarrow Prob(CR) = \alpha$$

A posteriori significance level (p-value):

The true significance level of a statistical test is given by the *p-value* $= \alpha^*$, and it represents the minimum significance value at which the null hypothesis begins to be rejected. Thus, for any significance level established a priori $\alpha \geq \alpha^*$, there will be sufficient statistical evidence to reject H_0 $(\theta_1 = \theta_2)$ and accept H_1 $(\theta_1 \neq \theta_2)$. The a priori significance value of the ESI is fixed at $\alpha = 0.05$, so we can reject H_0 when the p-value < 0.05. The p-value can be calculated as follows:

$$\alpha^* = Prob\left(\left| \hat{\theta}_1 - \hat{\theta}_2 \right| > \left| \hat{\theta}_1 - \hat{\theta}_2 \right|_{obs} \right) = Prob\left(\hat{\theta}_1 - \hat{\theta}_2 < -\left| \hat{\theta}_1 - \hat{\theta}_2 \right|_{obs} \; \text{ó} \; \hat{\theta}_1 - \hat{\theta}_2 > \left| \hat{\theta}_1 - \hat{\theta}_2 \right|_{obs} \right)$$

Given the symmetry of the t-distribution, the p-value can also be calculated as follows:

$$\alpha^* = 2 \cdot Prob\left(\left| \hat{\theta}_1 - \hat{\theta}_2 \right| > \left| \hat{\theta}_1 - \hat{\theta}_2 \right|_{obs} \right)$$

The observed standardized difference is as follows:

$$\alpha^* = Prob(|t| > |t_{obs}|) = 2 \cdot Prob(t > |t_{obs}|)$$

- **Evaluation of estimate quality[38]**

During the analysis of results, estimate quality is evaluated for tabulations subject to publication. The standard consists of two stages: The first evaluates whether each of the estimates in a tabulation meets the quality criteria for categorization within the same tabulation. The second evaluates whether the tabulation consists of a minimum number of estimates that meet the quality criteria in order to determine whether the tabulation, as a whole, can be published.

---

[38] For more information, see *"Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares"* (Standard for evaluation of estimate quality in household surveys) at https://www.ine.cl/docs/default-source/institucionalidad/buenas-pr%C3%A1cticas/clasificaciones-y-estandares/est%C3%A1ndar-evaluaci%C3%B3n-de-calidad-de-estimaciones-publicaci%C3%B3n-27022020.pdf.

1. First stage: flowchart for the evaluation of estimates

The flowchart consists of a set of criteria[39] and decisions for the evaluation of the estimates in the tabulations that are subject to publication. Each criterion presents the tolerance range of values for the estimate, and the proposed decisions must be made according to whether the estimate is within this range.

Three criteria are involved in the flowchart (Figure 2): the first two refer to the sample size and the degrees of freedom from which the estimates are elaborated; the third criterion, depending on the type of estimator, evaluates the estimate according to the standard error or coefficient of variation.

Figure 2 Flowchart for the evaluation of estimate quality

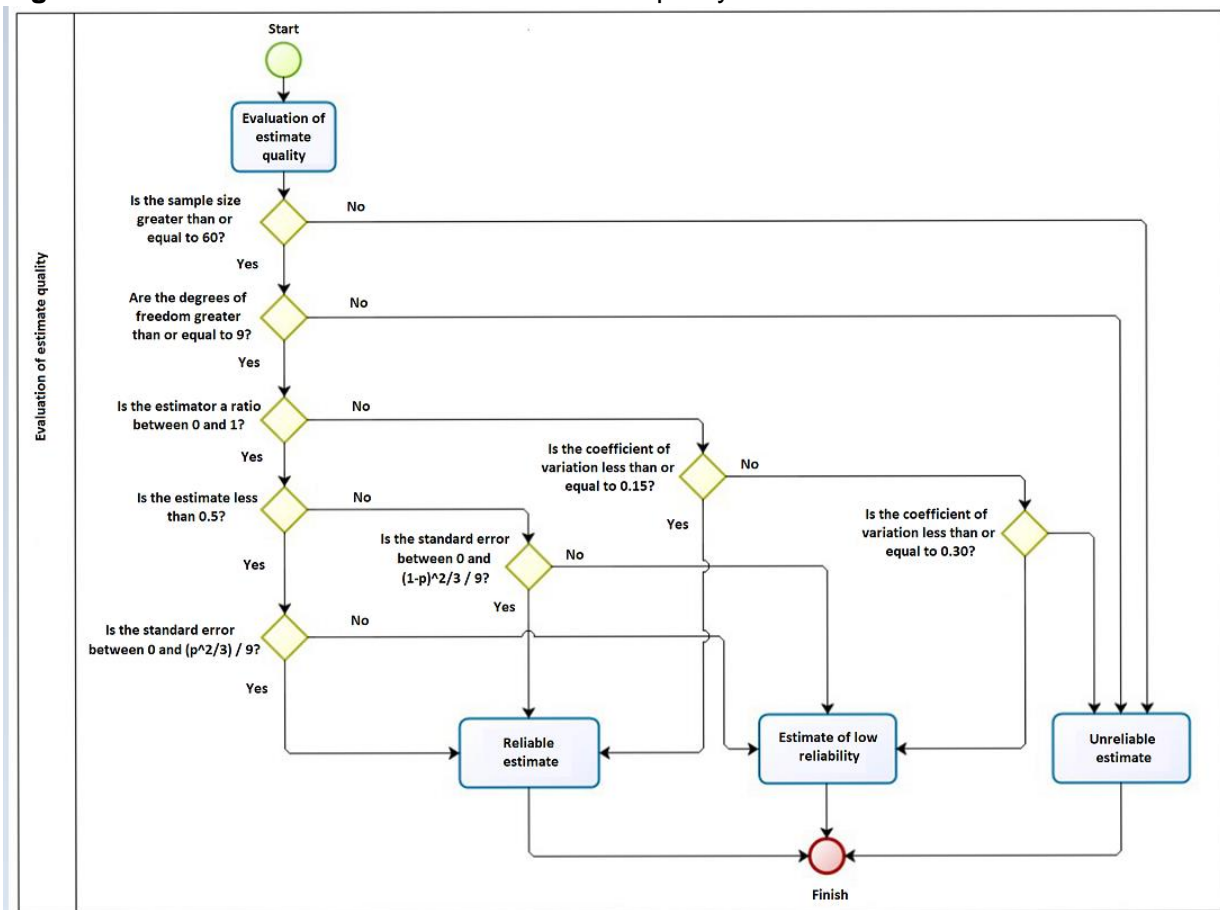**Source:** "*Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares*" (Standard for evaluation of estimate quality in household surveys), INE 2020.

---

[39] The criteria evaluated in the flowchart are applied sequentially. Thus, if an estimate does not meet any of the criteria in the flowchart, it does not merit evaluation in the subsequent stage.

According to the decision flowchart defined in the standard, estimates can be divided into three types: *reliable estimates, estimates of low reliability,* and *unreliable estimates*.

(a) *Reliable estimates* meet all the proposed quality criteria, and they can therefore be recognized as reliable and accurate estimates of a population parameter.

(b) *Estimates of low reliability* meet the criteria for sample size and degrees of freedom, but they do not meet the requirements established for the standard error or coefficient of variation, so the estimate should be used with caution when interpreting population characteristics because it could lead to inaccurate conclusions. Specifically, this category includes estimates where the standard error does not meet the acceptance threshold and estimates whose coefficient of variation is greater than 15% and less than or equal to 30%.

(c) *Unreliable estimates* do not meet either of the first two criteria established for the sample size or the degrees of freedom. Also included in this category are estimates that are not proportions or ratios between 0 and 1 whose coefficient of variation exceeds 30%. In these cases, the use of the estimates is not recommended.

2. Second stage: quality criteria for evaluation of the tabulation

Once it has been determined whether the estimates for each tabulation are "reliable", "of low reliability" or "unreliable", it must be decided whether to publish each tabulation. This process corresponds to the second stage of quality criteria evaluation, as illustrated in the flowchart in Figure 2.

The flowchart first requires accounting for the number of estimates in tabular form. For example: Table 4 has three tabulations, the first is yellow, the second green, and the third blue. Then, the percentage of reliable estimates is calculated. If more than 50% of tabulation estimates are classified as "reliable", it can be published as a graph.

Table 4 Reliability of the estimates of income from main job and percentage of employed, according to ICSE

ICSE	Mean	Median	Percentage of employed
Employer	1: Unreliable	1: Unreliable	1: Unreliable
Own-account workers	2: Reliable	2: Reliable	2: Reliable
Paid employees of the private sector	3: Reliable	3: Reliable	3: Reliable
Paid employees of the public sector	4: Reliable	4: Reliable	4: Reliable
Domestic workers	5: Unreliable	5: Unreliable	5: Unreliable

Source: Supplementary Survey of Income (ESI)

Figure 3 shows a flowchart that summarizes the sequence for evaluating tabulations.

Figure 3 Flowchart for evaluating the quality of estimates
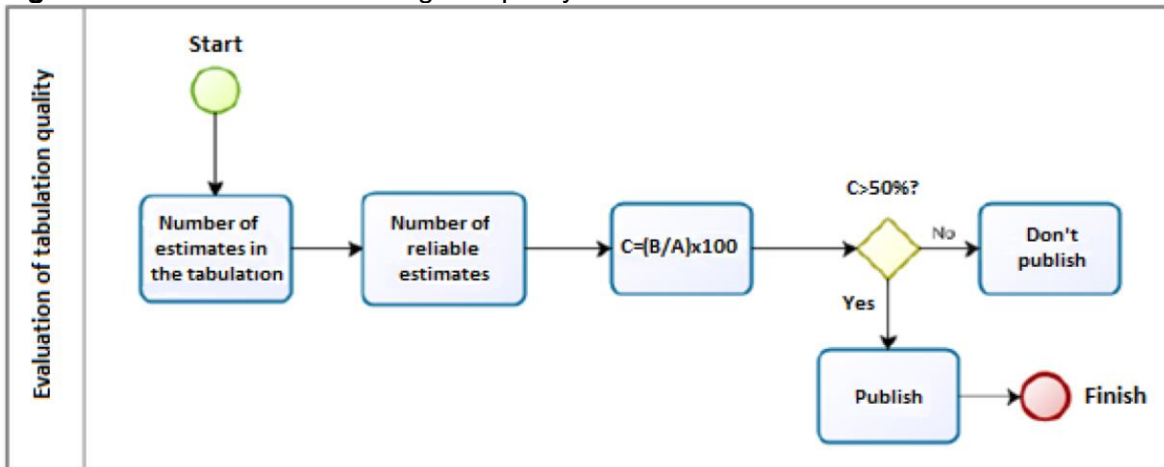
Source: "Estándar para la evaluación de la calidad de las estimaciones en encuestas de hogares" (Standard for evaluation of estimate quality in household surveys), INE 2020.

4.7.2   Expansion factor

The methodology for calculating the expansion factor did not change with respect to the 2019 ESI. The explanation of the expansion factor can be found at www.ine.cl, section *Ingresos y Gastos de las personas*, *"Documento Metodológico Factores de Expansión - ESI 2019"*[40].

4.8 Estimators

4.8.1   Horvitz-Thompson estimator

The Horvitz-Thompson estimator is obtained by multiplying the variable of interest by the expansion factor associated with each observation (Horvitz et al, 1952), thus approaching the projected population number. Mathematically, it can be expressed as follows:

$$\hat{\pi}_{Fy} = \sum_{i=1}^{n} F_i \cdot y_i$$

Where

F_i: ESI expansion factor associated with person i.

y_i: Variable of interest associated with person i. For example, income from main job. To calculate the number of employed persons, the variable y_i will take the value of 1 if the person i belongs to the population under analysis, and 0 if not.

---

[40] For more information, see Documento Metodológico Factores de Expansión - ESI 2019 (Methodological Document, Expansion Factors - 2019 ESI).

4.8.2 Ratio estimator

The ratio estimator is the quotient between the Horvitz-Thompson estimator of a variable of interest (x_i) over the Horvitz-Thompson estimator of another variable of interest (y_i). Mathematically, it can be expressed as follows:

$$\hat{R} = \frac{\sum_{i=1}^{n} F_i \cdot x_i}{\sum_{i=1}^{n} F_i \cdot y_i} = \frac{\hat{\pi}_{Fx}}{\hat{\pi}_{Fy}}$$

Where

F_i: ESI expansion factor associated with person i.

x_i: Variable of interest associated with person i. For example, for income from main job

y_i: Variable of interest associated with person i. The example measures the number of hours worked.

The following are more examples:

- Hourly income: the total income from the main job of paid employees ($\hat{\pi}_{Fx}$) divided by the total monthly hours worked in the case of paid employees ($\hat{\pi}_{Fy}$).

- This expression is also used when the variable of interest (y_i) is binary (0,1), which is known as a proportion. For example, the proportion of paid employees is the total number of paid employees who worked for at least one month ($\hat{\tau}_{Fx}$) divided by the total number of employed persons who worked for at least one month ($\hat{\tau}_{Fy}$).

$$\hat{R} = \frac{\hat{\tau}_{Fx}}{\hat{\tau}_{Fy}}$$

4.8.3 Estimator for the averages

The estimator for averages is obtained from the quotient between the Horvitz-Thompson estimator of a variable of interest (x_i) over the Horvitz-Thompson estimator of the units that present information on the variable of interest. This estimator could be interpreted as a special case of the ratio estimator. Mathematically, it can be expressed as follows:

$$\widehat{mean} = \frac{\sum_{i=1}^{n} F_i \cdot x_i}{\sum_{i=1}^{n} F_i \cdot z_i} = \frac{\hat{\lambda}_{Fx}}{\hat{\lambda}_{Fz}}$$

Where

F_i: ESI expansion factor associated with person i.

x_i: Variable of interest associated with person i. For example, the income from the main job.

z_i: Variable that will take the value 1 when person i presents information for the variable of interest (x_i), and 0 otherwise.

For example, the average income of the main job is the total declared income ($\hat{\lambda}_{Fx}$) divided by the total number of persons in the population ($\hat{\lambda}_{Fz}$)).

V. Methodology for data collection and processing

5.1 Techniques and methods of data collection

The information is collected between October and December of each year. The data reference period is the month prior to the reference week, which can be September, October, or November.

Data is collected with a mobile device (tablet) during face-to-face interviews, through a process known as Computer-Assisted Personal Interviewing (CAPI).

Among the main measures taken during the COVID-19 pandemic was the initial suspension of face-to-face collection in favor of telephone collection in which an INE interviewer contacts the household remotely in order to complete the survey.

5.2 Training method

Training focuses on reinforcement of content and the introduction of new concepts. The development of the support material involves providing instructions on any modifications to the questions, if necessary, and creating case studies that simulate situations.

In 2021, training was intended for all interviewers, supervisors, and coordinators, and it was conducted through videoconferences in September. The materials included the following:

- A presentation with case studies
- Interviewer manual
- Manual of specific cases

5.3 Analysis for data improvement.

- **Cleansing**

Survey cleansing

Survey cleansing is an analysis of the consistency of data, which is based on the income profile of each person or household. Both labour and non-labour income are cleansed, as are the variables of hours and days worked. Once the central value of the distribution and its dispersion are taken into account, outliers among records are identified according to relevant labour and sociodemographic characteristics. Thus, persons and households are selected for further review and, if necessary, cleansing.

Cleansing in the office

By cleansing in the office, conceptual errors can easily be identified and remedied through the use of software. The following are examples of data that have been cleansed in the office:

- Employed persons classified as providing live-in domestic service who report receiving income in the form of food and housing benefits: the amounts or percentages reported in the items of housing benefits (question d3_1) and food benefits (question d3_2) are eliminated for live-in domestic service because these benefits are necessary for the performance of their work because these workers live at their place of employment.

Cleansing with RStudio and Stata software

Cleansing with these software packages does not necessarily result in an absolute correction that transforms "false" data into "true" data (the true value is not known), rather it aims to establish a consistent framework in logical, mathematical, and statistical terms to which the data must conform. Data is cleansed by implementing a requirements system, a set of logical controls to detect inconsistencies and potentially erroneous data.

- **Deflation**

The nominal information in pesos published each year by the Supplementary Survey of Income (ESI) is anchored to stable values that allow for comparison over time. These anchors are called real values (or constant prices), and their purpose is to discount the effect of inflation on variables valued in current pesos, as is the case of monetary income reported by persons in a survey such as the ESI.

Because current income of a given month is not directly comparable with that of any other month, it must first be converted to constant pesos of a common period (deflation) to make it comparable.

Because the estimates are constructed by aggregating the figures for the quarter as a whole, it is essential that the microdata be expressed in the same monetary units, which in this case are pesos for the month of October of each year.

1. **Deflation of nominal values in the ESI.**

Deflation consists of multiplying nominal values by a price index (or deflator). For example, we can deflate a person's nominal income.

$$Nominal\ income \cdot deflator = Real\ income$$

The deflators calculated in the ESI are constructed using the CPI[41] with pesos valued for October, with the base year 2018=100[42].

Because the survey collects monthly figures and annual averages for some items[43], the algorithms had to be differentiated for the construction of a monthly and an annual deflator.

2. Methodology for the Creation of Monthly and Annual Deflators

2.1 Monthly figures

October of each year is taken as the central month, bearing in mind that the ESI information is collected according to the calendar shown in Table 5:

Table 5 Correspondence between the survey period and the reference month for income.

Month of collection	Reference month for income
October	September
November	October
December	November

Table 6 refers to the formulation of each deflator and its respective value for the year 2021. These values are used to adjust the income data to October:

Table 6 Month of survey, monthly deflator, and mathematical formula, 2021 ESI.

Month of collection	Formula	Deflator
October	$\left(\dfrac{CPI_{October}}{CPI_{September}}\right)$	1.0133692238672
November	$\left(\dfrac{CPI_{October}}{CPI_{October}}\right)$	1.000000000000000
December	$\left(\dfrac{CPI_{October}}{CPI_{November}}\right)$	0.99497841599859

In summary, the values collected in October with respect to the reference month of September were multiplied by a factor of 1.0133692238672 to obtain the income in real terms at October prices.

---

[41] INE publishes the Consumer Price Index every month.

[42] The latest bulletin in English translation can be found at https://www.ine.gob.cl/statistics/economic/price-indices-and-inflation/consumer-price-index.

[43] Certain self-employment income, as well as cash flows from financial assets, use annual averages.

2.2 Annual averages:

The methodology for obtaining the annual-average deflators was slightly modified from the methodology used in the first version of the ESI (2010). As of 2011, the methodology has been to use an arithmetic average of the monthly CPI variations of the year prior to the income reference month, which required adjustment factors that eliminated the effects of possible disproportionate changes in the price index from one year to the next (i.e., sudden rises or falls in the price index). The result of this process provided greater stability to annual income by smoothing the changes[44].

The calculation is as follows:

- Annual-average deflator in October.

$$\left[\sum_{t=Oct\ 2020}^{Sept\ 2021} \left(CPI_t / CPI_{t-1} \right) \right] * \frac{1}{12}$$

- Annual-average deflator in November.

$$\left[\sum_{t=Nov\ 2020}^{Oct\ 2021} \left(CPI_t / CPI_{t-1} \right) \right] * \frac{1}{12}$$

- Annual-average deflator in December.

$$\left[\sum_{t=Dec\ 2020}^{Nov\ 2021} \left(CPI_t / CPI_{t-1} \right) \right] * \frac{1}{12}$$

Table 7 shows the resulting deflators for 2021.

---

[44] For annual averages, deflators were constructed by dividing the CPI for the month of October by the monthly average of the index for the last 12 months.

Table 7 Month of survey and annual deflator 2021 ESI.

Month of collection	Annual deflator
September	1.00435060993764
October	1.0048976041761
November	1.00542770707147

The deflators for monthly and annual averages used in previous years are shown in the following table:

Table 8 Month of survey, monthly and annual deflator, according to ESI 2010-2021.

Version	Reference month for income	Monthly figures	Annual averages
2010 NESI	September	1.00097866510080	1.00155414108277
	October	1.00000000000000	1.00164400491595
	November	0.99931607230093	1.00203340735920
2011 NESI	September	1.00483320697498	1.00268587130787
	October	1.00000000000000	1.00300708313072
	November	0.99680361004043	1.00234292226246
2012 NESI	September	1.00562108367121	1.00240857865381
	October	1.00000000000000	1.00176718720718
	November	1.00451030927835	1.00162717626789
2013 NESI	September	1.00135562584727	1.00127172144923
	October	1.00000000000000	1.00196177740320
	November	0.99622370077324	1.00418913310216
2014 NESI	September	1.01042258859200	1.00418913310216
	October	1.00000000000000	1.00493315535332
	November	0.99981248828052	1.00464207647846
2015 ESI	September	1.00407461064832	1.00379546553503
	October	1.00000000000000	1.00326646737305
	November	1.00027061158218	1.00322829360155
2016 ESI	September	1.00166871596698	1.00254596435695
	October	1.00000000000000	1.00234547313350
	November	0.99947419156953	1.00241185841865
2017 ESI	September	1.00588693619600	1.00120274548439
	October	1.00000000000000	1.00155426383680
	November	0.99914008083240	1.00158214502135
2018 ESI	September	1.00360919926137	1.00258307143562

Version	Reference month for income	Monthly figures	Annual averages
	October	1.00000000000000	1.00239326002440
	November	1.00000000000000	1.00232153841899
2019 ESI	September	1.00818474130371	1.00185100701764
	October	1.00000000000000	1.00223420447735
	November	0.99922742636408	1.00229863539130
2020 ESI	September	1.006805293005670	1.00254231039284
	October	1.00000000000000	1.00242735636801
	November	1.00131603684903	1.00225339985636
2021 ESI	September	1.0133692238672	1.00435060993764
	October	1.0000000000000	1.0048976041761
	November	0.99497841599859	1.00542770707147

3. Deflation of nominal values in the ESI

Since 2016, the ESI has incorporated year-on-year comparisons in real terms, both for labour income and for household income. Starting in 2021, comparisons will be only for labour income because questions on household income were eliminated.

A year-on-year analysis in real terms requires deflating each year's income to October 2021 prices in order to keep personal purchasing power constant. The mathematical formula for the deflators needed to adjust revenues to October 2021 prices is presented below.

$$Deflator_{year\ i} = \frac{CPI_{October\ 2021}}{CPI_{October\ i}}, where\ i = \{2010, \dots, 2021\}$$

Table 9 shows the values of deflators for each year from 2010 onwards because this is the period in which the survey is comparable. However, the results from 2015 to 2021 are shown because it facilitates the graphical representation of the information.

Table 9 Year of survey and deflator 2021 ESI

Year of collection	Deflator
2010	1.4446
2011	1.3935
2012	1.3539
2013	1.3334
2014	1.2571
2015	1.2088
2016	1.1754

Year of collection	Deflator
2017	1.1537
2018	1.1211
2019	1.0915
2020	1.0603
2021	1.0000

- **Imputation**

Both total and partial non-response[45] are a recurrent situation in household surveys, especially in income surveys such as the ESI, and it can significantly impact estimates and their subsequent inference about the population. Non-response has various causes, the most relevant of which are respondent fatigue, lack of knowledge of the information requested, and refusal to answer the income survey.

The imputation process requires the evaluation of non-response rates on the income questions. In general, the gross non-response rates for each year in the former ESI (which continue to be measured in the current ESI) are within acceptable ranges for this type of survey (under 10%). This range can be explained by operational and technical factors, including quality control of data collection and the supervision instruments implemented to minimize errors. The gross partial non-response rates calculated include cases with refusal (*no response*) and cases of non-response due to lack of information (*not sure*)[46].

The items subject to omission that are considered for imputation are divided into two main groups: (a) labour Income and (b) hours actually worked.

The following cases, according to variable, are imputed:

 (a) Labour income. Persons who are classified as employed during the reference month in a category other than unpaid family members and who do not answer the questions on labour income. These items may be *wages and salaries* (income from dependent work) or *self-employment income* (income of own-account workers or employers).

 (b) Hours actually worked. Persons who are classified as dependent employed who do not respond to questions on the number of hours or days actually worked in the reference month.

---

[45] The direct association between income survey and the employment survey means that there are no problems of TOTAL non-response, because non-response can be corrected through *an adjustment to the theoretical expansion factors* by eliminating the dwellings with no operational response. Thus, the remaining units can be used to estimate (without bias) the total population.

[46] See Annex 3 for the non-response rates.

To impute values for these items, a procedure to reduce possible estimation biases must be used. All the available information provided by the survey itself must be used, whether the labour characteristics of all income recipients who reported their income or persons' sociodemographic variables.

For the imputation of labour income, the method used for the ESI is a variant of the **conditional means method** (Medina et al, 2007; Keifman et al, 1998), which was especially designed for the survey. This method is applied at the level of income recipients, and it is based on the search for a group of persons who reported income *(donors)* and who have characteristics similar to those of the recipient. The *median* income of this group is thus imputed[47].

One of the advantages of this procedure is that it minimizes bias from measurement errors because the search seeks donors with similar characteristics to the recipient. In addition, geographical variables are incorporated to account for differences in jobs and income in different parts of the country.

The donor search is based on two fundamental criteria:

i.   *The establishment of minimum requirements* uses a set of categorical variables or labour and sociodemographic characteristics of employed persons. This method uses a series of minimum requirements that gradually make the variables more flexible.

ii.   *The use of geographical estimation areas*, ranging from the most disaggregated level possible (conglomerate) to the most aggregated level (country).

As a result, the procedure combines both qualitative variables (minimum requirements based on the characteristics of income earners) and geographical variables (estimation areas). Geographical estimation areas are applied vertically by going through all the estimation areas before moving on to the next level of the requirement in order to favor donors who have the same characteristics as the recipient rather than donors who have fewer characteristics in common.

The methodology for imputing hours actually worked is the imputation by **substitution**. According to this method, data from the same observation are used for the imputation. For hours worked, the hours usually worked during the reference week (which are answered in the employment module) are used[48].

---

[47] For asymmetric distributions, the median is a better estimator than the mean because it is not affected by the extremes of the distribution (tails).

[48] For more information, see Annex 4.

The main advantage of this methodology is that it uses information from the same observation units, so the labour and sociodemographic profile of the "donor" and the recipient correspond. For this reason, it is usually recommended in the literature as the first imputation option[49].

---

[49] This topic was already mentioned in the section "Milestones and major changes 2010–2021" of this document.

VI. Presentation of results

6.1 Dissemination of results

Between 2010 and 2017, the delivery of ESI results used the following formats:

- tables of estimates and coefficients of variation
- summary of national and regional results

From the publication of the 2018 ESI to the 2020 ESI, the tables of estimates and coefficients of variation were replaced by twenty-one income indicators made available at INE.STAT. Since the publication of the 2021 ESI, only eighteen of the twenty-one indicators have been updated because the other three involve household income, which is no longer measured.

The summary of national and metropolitan-regional results is available on the INE website, under the section *Ingresos y Gastos de las personas*. Regional results are available on websites of the respective regions and on the national microsite.

6.2 Microdata

At the personal and household levels, ESI databases are available in an external version for the years 2010 to 2020. These databases provide information at the personal or household level, but they provide them as a subset of variables in order to safeguard the confidentiality of the persons who answered the survey. As a result of the elimination of questions D10 to D14 (household income questions), only the external version of the ESI database of persons is available for the 2021 ESI and later versions.

All databases contain cleansed, deflated, and imputed data. The databases of persons are cleansed in the office, income variables are deflated, and they are imputed if the employed persons did not report income from their main job. For household databases, in addition to the procedures mentioned above, values are imputed when heads of household did not report the type of dwelling and/or imputed rent, when elderly persons did not report the amount in retirement benefits and pensions, or when dependent employed persons did not report hours and/or days actually worked.

The databases are available (in Spanish) to the public at https://www.ine.gob.cl/estadisticas/sociales/ingresos-y-gastos.

6.3 Metadata

Metadata provides information about the data produced, including description and characterization, in order to provide a better description. The metadata produced for the ESI 2021 are mentioned below:

- Technical specifications sheet

- Questionnaire
- Variable guidelines for the databases of persons
- Statistical note
- Methodological document
- Methodological document on expansion factors
- Supplementary technical note

These metadata are available on INE's website at the ESI page under the section *Ingresos y Gastos de las personas*.

6.4 Legal framework and database accessibility

- **Copyright law**

• Law No. 17.336 on Intellectual Property, Article 88, first paragraph: "The State, municipalities, official corporations, semi-fiscal and autonomous institutions, and other legal persons of the State will be owners of copyright in regard to works produced by its officials in the course of their duties".

- **Data Privacy Statement**

Excerpt from Organic Law 17.374, which created the National Statistics Institute, published on 10-12-1970:

• Article 29: "The National Statistics Institute, fiscal and semi-fiscal bodies, State enterprises, and each of their respective officials, may not disclose facts that refer to specific persons or entities that they have become aware of in the course of their activities. The strict maintenance of these facts constitutes "statistical secrecy". Its violation by any person subject to this obligation will result in the charge provided for in article 247 of the Criminal Code, and custodial sentencing must be applied in all cases".

Article 30: "Statistical data may not be published or disseminated with express reference to the persons or entities to whom they directly or indirectly refer, if prohibited by the person or persons concerned".

- **Database accessibility**

As of the publication of the ESI 2016, the databases from 2010 onwards have been available in Spanish to the general public without prior request at the institutional website[50].

---

[50] For more information, see the ESI web page (in Spanish) at https://www.ine.cl/estadisticas/sociales/ingresos-y-gastos/encuesta-suplementaria-de-ingresos.

VII. Key concepts

This section describes the socioeconomic and sociodemographic variables considered in the design, execution, and delivery of the statistical results. The following is a definition of concepts associated with the tabulations of persons.

- **Sociodemographic variables**

Number of employed persons: total number of employed persons whose current job (in the reference week) is the same as their job in the reference month. The data are expanded using the expansion factor for persons (fact_cal_esi), which is adjusted to total population projections[51].

Number of dependent employed persons: total number of employed persons whose current job is the same as the one they had in the reference month, and who are dependent and/or subordinate to an employer, enterprise, business, or institution where they work. This group includes public and private paid employees and live-in and live-out domestic workers.

Number of paid employees: total number of employed persons whose current job is the same as the one they had in the reference month and who work as paid employees. This variable is a subgroup of dependent employed persons that includes public and private paid employees.

Number of own-account workers: total number of employed persons whose current job is the same as the one they had in the reference month, who work on their own account or with one or more partners, and who have not hired any paid employee on a continuous basis to work for them during the reference period (partners are not necessarily members of the same family or household).

- **Income variables[52] of persons**

Total income from main job: total income from current main activity received in the reference month of the survey.

The labour income of dependent workers includes wages and net monthly salary, regular variable income, and benefits in cash and in kind. The labour income of self-employed workers includes assigned salaries, net earnings, withdrawals from own business, and own consumption.

---

[51] This is one of the reasons why the number of employed persons for the same calendar quarter differs between the ENE and the ESI.

[52] All income figures are expanded using the expansion factor for persons (fact_cal_esi).

Regular variable income: all variable income provided by the employer at least once a year and actually received in the reference month. Included are productivity bonuses, other bonuses, sales commissions, overtime pay, and other variable income.

Benefits in cash and in kind: all benefits provided by the employer that are associated with goods and services for the worker's use. These benefits may be in cash or in kind, and they include housing, food, transportation, and gasoline, among other benefits.

Own-consumption: the part of the production (of goods and services) of the enterprise, business, or independent activity that is intended for own or household consumption.

Income from other jobs: all income received during the reference month from jobs other than the current job or situation. This income may be in cash or in kind.

Mean income: the sum of all income of the population, divided by the total number of individuals or households receiving that income.

Median income: the income received by the individual or household located in the middle of the distribution, after ordering income from lowest to highest.

Net minimum income: the amount of the legally mandated minimum gross income[53] minus the legal discounts for social security and health, which added together represent approximately 19.5% of minimum gross income.

---

[53] By law, the minimum income is published each year in the Official Gazette.

VIII. Bibliography

- Grupo Intersecretarial de Trabajo sobre Cuentas Nacionales (GITCN) (1993). Sistema de Cuentas Nacionales, 1993. Accessed at http://comuna.cat/-/sctasnac93.pdf.

- Grupo Intersecretarial de Trabajo sobre Cuentas Nacionales (GITCN) (2009). Sistema de Cuentas Nacionales, 2008. Accessed at https://unstats.un.org/unsd/nationalaccount/docs/SNA2008Spanish.pdf.

- Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association, 47(260), 663-685.

- Instituto Nacional de Estadística (2016), CAENES Clasificador de Actividades Económicas Nacional para Encuestas Sociodemográficas. Accessed at http://www.ine.cl/estadisticas/laborales/ene.

- Instituto Nacional de Estadística (2018), Manual conceptual y metodológico, diseño muestral, Encuesta Nacional de Empleo (ENE). Santiago. Accessed at https://www.ine.cl/docs/default-source/laborales/ene/publicaciones/clasificador-de-actividades-econ%C3%B3micas-nacional-para-encuestas-sociodemogr%C3%A1ficas-(caenes).pdf.

- Keifman, S. (1998). Imputación de ingresos de hogares: la experiencia de la Encuesta Nacional de Gastos de los Hogares de la Argentina. En: Segundo Taller Regional sobre Medición del Ingreso en las Encuestas de Hogares-LC/R. 1886-1998-p. 423-430.

- Medina, F., & Galván, M. (2007). Estudios estadísticos y prospectivos. Imputación de datos: Teoría y práctica. División Estadística y proyecciones Económicas Naciones Unidas. CEPAL.

- Ministerio de Desarrollo Social. Documento metodológico CASEN. Diversos años.

- OIT (1993), Resolución sobre la Clasificación Internacional de la Situación en el Empleo (CISE), adoptada por la 15° Conferencia Internacional de Estadísticos del Trabajo, Ginebra. Accessed at http://www.ilo.org/public/spanish/bureau/stat/download/res/icse.pdf.

- OIT (1998), Resolución sobre la medición de los ingresos relacionados con el empleo, adoptada por la 16ª Conferencia Internacional de Estadísticos del Trabajo. Ginebra. Accessed at http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_087492.pdf.

- OIT (2003), Informe de la Conferencia, 17° Conferencia Internacional de Estadísticos del Trabajo. Ginebra. Accessed at http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/meetingdocument/wcms_087590.pdf.

- OIT (2004), ¿Qué es el trabajo decente? Accessed at http://www.ilo.org/americas/sala-de-prensa/WCMS_LIM_653_SP/lang--es/index.htm.

- OIT (2013), Resolución I, Resolución sobre las estadísticas del trabajo, la ocupación y la subutilización de la fuerza de trabajo de la 19° Conferencia Internacional de Estadísticos del Trabajo. Ginebra. Accessed at http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_234036.pdf

- The Canberra Group (2001), Expert Group on Household Income Statistics, Final Report and Recommendations, Ottawa. Accessed at http://www.lisdatacenter.org/wp-content/uploads/canberra_report.pdf

- UNESCO (1997), Organización de las Naciones Unidas para la educación, la ciencia y la cultura. Clasificación Internacional Normalizada de la Educación CINE1997 Accessed at http://unesdoc.unesco.org/images/0014/001470/147002s.pdf.

- United Nations (2005), Informes Estadísticos Serie M, N° 4, Rev.3.1. Clasificación Industrial Internacional Uniforme de todas las actividades económicas (CIIU), revisión 3.1. Accessed at https://unstats.un.org/unsd/publication/SeriesM/seriesm_4rev3_1s.pdf.

- United Nations (2011), United Nations Economic Commission for Europe. Canberra Group, Handbook on Household Income Statistics, Second Edition. Geneva.

ANNEXES

Annex 1. Comparability of educational transfers

1. Measurement of educational transfers

Since its beginnings, the ESI has had a question on transfers received as **"scholarships"**. However, in the previous measurement (before 2010), this information was part of a larger item that included income from the *single family allowance, welfare pension, and other State subsidies*, and therefore specific information on scholarships could not be disaggregated.

The question on scholarships was as follows:

> **6** Did you receive income from the single family allowance, pension assistance, scholarships, or other state subsidies?
>
> Yes ☐   No ☐ (Go to Question 7)
>
> **6.1** How much did you receive in the previous month? $ ....

With the improvements to the income questionnaire in 2010, refinements were made to the items of current transfers and monetary subsidies, which clarified the effect of each of the components, including "scholarships". The question on scholarships that was part of the questionnaire until ESI 2020 was formulated as follows:

SPECIAL SUBSIDIES AND SCHOLARSHIPS					
D14_9 Unemployment and severance Insurance					
D14_10 Single family allowance					
D14_11a Scholarships as cash transfers					
D14_11b Scholarships as transfers of goods and services					
D14_12 Other State subsidies					

The item "scholarships" includes both cash transfers freely available to the beneficiary as well as subsidies in cash and in kind that pass through a third party (i.e., through an educational institution) and that ultimately benefit the person. This group of benefits mainly consists of *scholarships* that the State transfers to public educational institutions or institutions with shared financing. The transfers consist of *food scholarships* and *scholarships that subsidize all or part of the educational fee*, at all educational levels, including the university level.

In addition to the change to the questionnaire, the use of a reference table was incorporated into field work in order to avoid the underestimation of the figures declared. Learning how to use the reference table resulted in a considerable increase in the number of "scholarships" captured in 2011 compared to 2010, as well as an increase of 76.4% in the number of people who reported the benefit in 2011 compared to 2010.

An analysis of the sample distribution of the figures captured in the ESI 2010 and 2011 shows that for the central-tendency indicators the mean number of scholarships reported varied 27.9% while the median varied 40.6% (see Table 1). At the same time, the number of scholarships in all percentiles is the highest for decile 10 (82.7%) and the lowest for decile 90 (18.4%).

Since 2011, the learning process for capturing this information has been gradually and steadily established. As a result, when analyzing the sample distribution of the number of cases and mean over the first five years since the incorporation of the scholarship variable (i.e., between 2010 and 2014) there was a variation of 0.9% and -1.3% for the last two years in the number of people who reported having the benefit and a variation in the mean of 17.4% and 0.4%, respectively (see Table 1).

Table 1 Descriptive statistics and annual percentage variation of the variable "Scholarships", ESI 2010–2014.

Statistics		Estimate					Variation (%)			
		2010	2011	2012	2013	2014	2011	2012	2013	2014
No	Yes	10,011	17,656	18,372	18,531	18,284	76.4%	4.1%	0.9%	-1.3%
	No	78,593	68,717	65,873	64,901	65,636	-12.6%	-4.1%	-1.5%	1.1%
	NS/NR	185	148	309	290	343	-20.0%	108.8%	-6.1%	18.3%
Mean		60,766	77,726	87,219	102,369	102,811	27.9%	12.2%	17.4%	0.4%
Median		48,046	67,545	81,089	96,608	88,639	40.6%	20.1%	19.1%	-8.2%
Stand. Dev.		53,659	45,897	54,878	75,128	61,099	-14.5%	19.6%	36.9%	-18.7%
Percentiles	10	19,986	36,507	38,949	45,000	50,068	82.7%	6.7%	15.5%	11.3%
	20	26,982	44,194	50,000	57,799	60,907	63.8%	13.1%	15.6%	5.4%
	30	31,978	50,242	62,640	68,774	72,487	57.1%	24.7%	9.8%	5.4%
	40	39,973	59,801	66,519	81,083	86,688	49.6%	11.2%	21.9%	6.9%
	50	48,046	67,545	81,089	96,608	88,639	40.6%	20.1%	19.1%	-8.2%
	60	57,352	77,649	81,722	97,291	99,468	35.4%	5.2%	19.1%	2.2%
	70	68,079	86,723	91,916	108,176	103,786	27.4%	6.0%	17.7%	-4.1%
	80	82,961	99,876	107,235	126,176	127,463	20.4%	7.4%	17.7%	1.0%
	90	120,000	142,104	160,943	191,571	177,669	18.4%	13.3%	19.0%	-7.3%

When analyzing the behavior of the other items included in *transfers* (retirement benefits, pensions, survivor's benefits, state subsidies, and national and international transfers), it has been found that the survey effect is within acceptable ranges from one year to another and subject to the non-sampling errors typical of a survey of this type (for 2011–2010, 2012–2011, 2013–2012 and 2014–2013). From the 2021 ESI onwards, it will not be possible to study *transfers* (retirement benefits, pensions, survivor's benefits, state subsidies, and national and international transfers), because the questions associated with this item will no longer be asked.
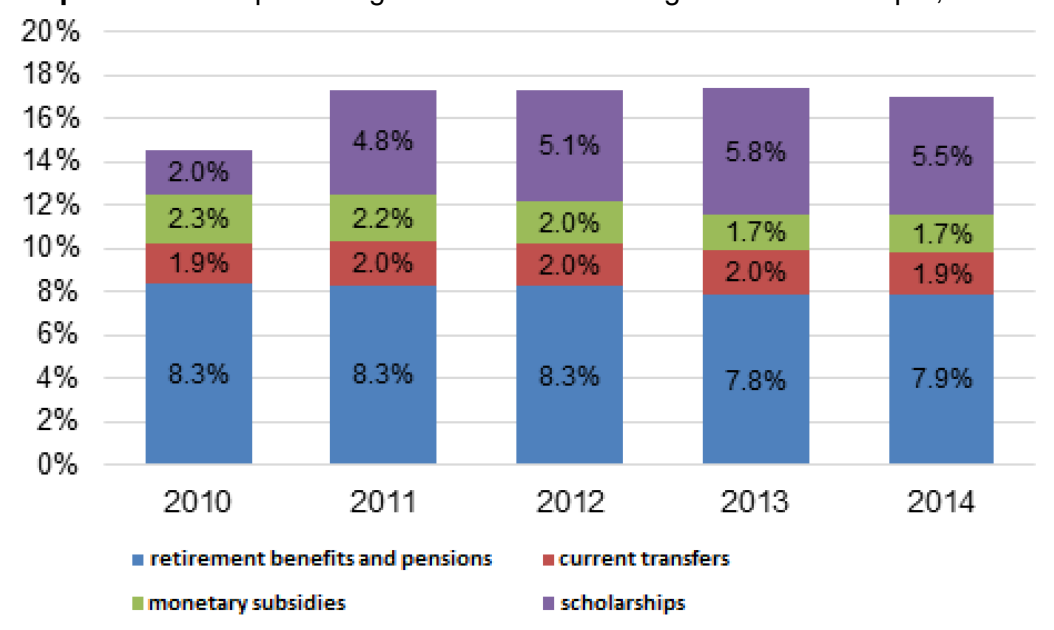
Effect on household-income structure

Therefore, if the item "scholarships" is included in household income, the income structure of households in 2011 *is not comparable* to that of 2010.

For 2011, the variation in total household income including "scholarships" was 13.9%, and the variation of transfers was 35.5%. Therefore, the share of this item in total income increased by 17.3% compared to 2010 (2.8 percentage points).

In large part, the jump in transfers in 2011 can be attributed to the "scholarship effect", which in turn was influenced by the increase in the share of scholarships as a part of total household income: from 2.0% in 2010 to 4.8% in 2011 (see Graph 1), and by the increase in their relative weight within transfers, from 13.9% in 2010 to 27.5% in 2011. However, as can be seen in Graph 1, the variations in the share of "scholarships" in total income were smaller and relatively similar from 2011 onwards. These variations were 0.3 pp for 2012–2011, 0.7 pp for 2013–2012, and 0.3 pp for 2014–2013, all of which suggest a stabilization in this income item within the total. The share of "scholarships" within transfers was 33.2% and 32.3% in the years 2013 and 2014, respectively, which is also consistent with the stabilization of this income item after the incorporation of the reference tables in the survey of 2011.

Graph 1 Structure percentage of transfers including item "scholarships", ESI 2010–2014.

It should be noted that the effect of the increased capture of "scholarships" has an impact on the analysis of household income only[54], specifically when aggregate income is constructed, because the analysis of personal income takes into account only the labour income received for work done in the previous month of the survey.

---

[54] These are for tabulations of Tables 1 to 5.

Annex 2. Programming the calculation of coefficients of variation (CV)

As mentioned in the chapter on sampling methodology, three main measurements of the estimation error are used in the ESI: the relative error, the absolute error, and the coefficient of variation. In fact, the statistical note published on the institutional website notes the relative and absolute errors of the main income estimates by region.

The ESI publishes the coefficient of variation for almost all of its estimators. Moreover, for the 2014 version, improvements were made to the calculation of this dispersion measurement, which is now done for complex samples entirely with the Stata and SPSS statistical packages.

Median income was added to the estimators of the 2015 ESI in order to complement the traditionally published indicators. Because of the complexity of CV calculation, which included a non-linear estimator for a survey with a complex design, the CVs of median income were not initially published. By concentrating institutional efforts, INE has published the CVs for median income since the 2016 ESI, and they are available to users. The CV for median income is also calculated with statistical packages for complex samples, but in this case with R software.

The main objective of this annex is to provide the user with the information for cross-referencing characteristics of the sample and calculating the CVs of the estimates, which in turn will be consistent with those calculated by the institution for its main tabulations.

It should be noted that the ESI has a stratified two-stage probabilistic sampling design whose strata variable is "*estrato*" in the database of persons and "*esth*" in the database of households. The first-stage selection units are blocks or sections[55], both of which are identified by the variable "*conglomerado*" (cluster)[56], and the second-stage selection units are dwellings, which are identified by the variable "id_identificacion" (in both databases). The expansion factors corresponding to each unit of analysis are "fact_cal_esi" (for the databases of persons and households).

The estimation method used by Stata and SPSS software is a linearization through the development of a Taylor series that approximates the corresponding analytical expressions[57]. The treatment given to strata with a single sampling unit is that of a known unit, and thus it does not contribute to the standard error.

The median CV uses the sub-bootstrap replication method with 2,000 replicates, which is done with R software for complex sampling designs. Replication methods are based on the idea of drawing a series of small samples from a large sample. In each sample replicate, the statistic of interest is calculated with the same original methodology. The variability of the

---

[55] These units are clusters of dwellings in urban areas (blocks) and rural areas (sections).

[56] For databases up to 2019, both units are identified as ***id_directory***.

[57] This is because, in complex multistage sampling designs, there may be no analytical expressions for calculating the variance.

estimate is used to derive the variance of the statistic of interest. Again, the strata with a single sample unit are treated as a known unit, and thus they do not contribute to the standard error.

Below are programming models in Stata and SPSS for calculating the coefficient of variation associated with estimates of mean income in the survey. First, the CVs are shown for estimates of persons and of households. An excerpt of the syntax used in R for the calculation of the CV of estimates for median personal income is also presented.

- **Programming**

Calculation of CVs of estimates of persons using Stata
Mean income by branch of economic activity

```
* Coefficient of variation of the mean income of the main job,
Level of branch of economic activity

Use "ESI_Personas_UsuariosExternos.dta", clear

svyset conglomerado58 [pweight = f_pers], strata(estrato) singleunit(certainty) || id_vivienda
svy, subpop(ocup_ref) : mean ( ing_t_p ), over(b14)

mat means = e(b)'
mat variances = vecdiag(e(V))'

local rows = `= rowsof(means)'
mat CV = J(`rows',1,.)
forval i = 1/`rows' {
        mat CV[`i',1] = 100*sqrt(variances[`i',1])/means[`i',1]
}

mat list CV
```

Calculation of median CV for estimates of persons, using R

Median income of employed persons by educational level, nationwide.

```
* Coefficient of variation of the median income of the main job
* By educational level (ISCED)


###Load database
esi_2020<- read_dta("C:/Users/Admin/esi_2020_personas_interna.dta")
```

---

[58] Use *id_directorio* for databases up to and including 2019.

###Create complex design
#To perform the repeated design in the ESI2020 database, through the #*as.svrepdesign* command, it is first necessary to use the variable *conglomerado_correlativo* in the #complex design statement. This procedure is done to avoid errors in the *survey* package associated with the character length of the original *conglomerado* (cluster) variable.

#For base years from 2010 to 2019, **id_directory** is used.
diseno <- svydesign(id = ~conglomerado_correlativo, strata = ~estrato,
                    weights = ~fact_cal_esi, data = esi_2020)

###Define the replicated design
set.seed(234262762)
diseno_rep <- as.svrepdesign(diseno, type = "subbootstrap",
            replicates = 2000)

###Generate data on median – se - cv - cv%:
mediana <- svyby(~ing_t_p, by = ~cine, data=esi_2020, na.rm.all=FALSE,
            subset(diseno_rep, ocup_ref==1),
            svyquantile, quantiles=c(0.5), method="constant",
            interval.type="probability", ties="discrete",
            vartype=c("ci", "se"), ci=TRUE)
Median

Annex 3. Non-response rates by income item, 2016–2021

Table 2 Partial non-response rates for income from paid jobs, 2016–2021(*)

NON-RESPONSE RATES IN THE SUPPLEMENTARY SURVEY OF INCOME, DEPENDENT EMPLOYED PERSONS, SERIES 2016–2021(*)

Area Estimation	Income from paid jobs					
	2016	2017	2018	2019	2020	2021
Region of Arica y Parinacota	2.10%	4.00%	3.00%	4.10%	12.55%	11.86%
Region of Tarapacá	0.20%	0.10%	0.60%	3.60%	1.80%	3.66%
Region of Antofagasta	6.50%	10.30%	8.00%	10.60%	6.67%	3.09%
Region of Atacama	3.20%	4.40%	3.40%	2.40%	4.39%	3.73%
Region of Coquimbo	2.60%	5.10%	4.50%	7.90%	7.61%	3.30%
Region of Valparaíso	1.90%	5.90%	7.10%	9.60%	6.30%	5.30%
Metropolitan region	14.40%	14.90%	15.80%	21.80%	19.12%	13.84%
Region of O'Higgins	1.50%	0.80%	2.80%	2.10%	4.34%	3.01%
Region of Maule	5.40%	7.00%	5.00%	8.30%	7.51%	5.61%
Region of Ñuble	5.30%	1.30%	0.80%	2.50%	6.08%	6.36%
Region of Biobío	2.00%	2.90%	1.20%	3.30%	2.34%	3.23%
Region of La Araucanía	3.10%	2.80%	2.90%	4.00%	4.95%	6.16%
Region of Los Ríos	1.10%	4.80%	4.80%	8.30%	11.20%	4.83%
Region of Los Lagos	3.90%	7.30%	15.20%	15.90%	14.43%	6.25%
Region of Aysén	2.00%	6.40%	2.30%	5.30%	5.62%	2.53%
Region of Magallanes y Antártica Chilena	0.00%	1.70%	0.20%	2.90%	3.52%	3.76%
Nationwide Total	**5.70%**	**7.20%**	**7.50%**	**10.10%**	**8.93%**	**6.39%**

Source: Supplementary Survey of Income, INE.

(*): These figures are based on the total of each reference subpopulation and on sample data from the October–December quarter of each year.

Table 3 Partial non-response rates for income from self-employment work, 2016–2021(*)

NON-RESPONSE RATES IN THE SUPPLEMENTARY SURVEY OF INCOME, SELF-EMPLOYED PERSONS, SERIES 2016–2021(*)

Area Estimation	Income from self-employment work					
	2016	2017	2018	2019	2020	2021
Region of Arica y Parinacota	1.50%	4.30%	2.90%	5.00%	11.96%	12.88%
Region of Tarapacá	0.20%	0.90%	0.20%	4.40%	2.24%	3.05%
Region of Antofagasta	6.20%	8.20%	3.70%	9.20%	4.57%	1.37%
Region of Atacama	3.60%	3.90%	2.20%	3.60%	8.57%	1.30%
Region of Coquimbo	1.50%	4.70%	4.70%	6.40%	5.63%	1.85%
Region of Valparaíso	2.80%	5.90%	6.20%	8.20%	7.28%	5.71%
Metropolitan region	13.60%	15.70%	14.00%	22.10%	18.33%	14.64%
Region of O'Higgins	2.30%	1.90%	3.00%	2.70%	5.39%	4.30%
Region of Maule	6.20%	9.00%	6.10%	11.00%	9.32%	7.15%
Region of Ñuble	4.30%	0.00%	1.30%	3.90%	7.87%	6.31%
Region of Biobío	1.90%	2.90%	1.80%	5.00%	2.87%	4.10%
Region of La Araucanía	3.80%	2.50%	3.10%	4.40%	6.69%	3.67%
Region of Los Ríos	0.60%	7.50%	6.50%	10.80%	10.93%	4.96%
Region of Los Lagos	2.90%	8.10%	10.00%	15.10%	11.25%	7.68%
Region of Aysén	2.40%	3.70%	2.30%	6.10%	3.92%	2.53%
Region of Magallanes y Antártica Chilena	0.00%	0.70%	1.50%	4.30%	10.48%	5.73%
Nationwide Total	**5.30%**	**7.40%**	**6.70%**	**10.50%**	**10.17%**	**6.73%**

Source: Supplementary Survey of Income, INE

Annex 4. Imputation methodology for hours actually worked (question D4)

In the publication of the results from the 2013 ESI, estimators related to the hourly income of paid employees were provided in response to the growing need for more and better estimators to improve the survey's characterization of household and employment income.

One of the required variables for the calculation of the estimator is from question D4, which asks only dependent employed persons about the days and hours they actually worked during the reference month.

Because question D4 allows respondents to indicate their lack of information or refusal, having to enter "888" or "999" to do so, the calculation for hourly income is reduced. These options are understood as *non-response* cases. Table 4 specifies for each year the number of sample observations that did not obtain a response to question D4, either in hours or days.

Table 4 NS/NR to question D4 in hours, days, or both 2010 – 2021

Data		Sampling observations		
		Hours (d4_horas)	Days (d4_dias)	Hours or Days
2021	Number	147	136	152
	%	0.15%	0.14%	0.15%
2020	Number	274	265	321
	%	1.47%	1.43%	1.73%
2019	Number	298	194	341
	%	1.03%	0.67%	1.18%
2018	Number	515	83	535
	%	1.64%	0.26%	1.70%
2017	Number	520	198	598
	%	1.64%	0.62%	1.89%
2016	Number	290	164	336
	%	0.93%	0.53%	1.08%
2015	Number	400	226	446
	%	1.27%	0.72%	1.41%
2014	Number	156	72	167
	%	0.48%	0.22%	0.52%
2013	Number	278	90	283
	%	0.87%	0.28%	0.89%
2012	Number	266	236	331
	%	0.82%	0.73%	1.03%
2011	Number	438	269	445
	%	1.35%	0.83%	1.37%
2010	Number	1067	492	1101
	%	3.23%	1.49%	3.33%

To overcome the information gap, a specific imputation methodology was designed for this question, which was applied until the 2014 production process, when the technical team in charge of the product made progress in analyzing and improving the methodology used, with the aim of delivering higher quality results.

The following is a detailed description of the imputation methodologies used for question D4, in addition to the calculation methodology for hourly income.

D4 imputation methodologies and calculation of hourly income

I. 2010–2013

Between 2010 and 2013, the methodology designed for the imputation of question D4 was based on calculating the median of the hours and days of the total number of cases that contained information, and then imputing this value for each of the missing observations or *non-responses*.

The advantages of this procedure are as follows:

1. By using the median of hours and days, the information obtained was not affected by extreme values as would have happened with the use of the mean.
2. The imputed information corresponded to a group of donors (dependent employed persons) that had greater homogeneity in the hours and days usually worked, which meant they were a good source of information for donation.

With this methodology, the following results were obtained:

Table 5 Hourly income for both sexes, female and male wage earners, and income gap (2010–2013)

Year	Hourly income Both sexes	Hourly income Women	Hourly income Men	Gender gap
2010	$2,183	$2,053	$2,256	-9.0%
2011	$2,364	$2,190	$2,463	-11.1%
2012	$2,679	$2,487	$2,795	-11.0%
2013	$2,863	$2,644	$2,998	-11.8%

II. 2014 – 2019

Responding to INE's commitment to continuous improvement and to the increasing importance of hourly income in the expansion in complementary information on labour income, the technical team of the survey conducted a series of analyses to improve the imputation process of question D4 in 2014.

From the analysis, it was determined that the best approximation for imputing the missing information for question D4 was to use the information on hours usually worked provided by the same respondent in module C (question C2) and adjust the figure to the reference

month[59]. The figure must be adjusted because the information comes from a module of the ENE, which uses the reference week instead of the reference month.

The basis for using this information is as follows:

1. The group subject to imputation, dependent employed persons, is more stable in the hours usually worked than is the self-employed group. Thus, we can assume that the hours usually worked during the reference week of dependent employed persons can be extrapolated to a month of work.

2. Until the 2021 ESI, question C2 was answered by all employed persons, so alternative information was available when information was missing for D4, and thus respondents' usual hours of work could be inferred from their work profiles because the same person was involved. Starting in 2020, respondents could declare "not sure" or "no response" to question C2 of the ENE. This option meant that if information was missing from D4, it was possible that the hours usually worked could not be imputed from other responses of the same respondent. In these cases, it was decided to apply the imputation method described in the following section.

With this methodology, the following results were obtained for hourly income for both sexes, female and male paid employees, and for the income gap:

Table 6 Hourly income for both sexes, female and male wage earners, and income gap (2014–2019)

Year	Hourly income Both sexes	Hourly income Women	Hourly income Men	Gender gap
2014	$3,023	$2,811	$3,146	-10.7%
2015	$3,203	$2,938	$3,358	-12.5%
2016	$3,356	$3,088	$3,512	-12.1%
2017	$3,565	$3,315	$3,718	-10.9%
2018	$3,817	$3,543	$3,990	-11.2%
2019	$3,889	$3,600	$4,068	-11.5%

III.   2020–2021

For the years 2020 and 2021, the same methodology was generally maintained as in the previous period. However, methodological decisions were made to account for the admission in variable C2 of partial non-response (not sure/no response) in these editions of the survey.

---

[59] If there is no information for question C2, the median of the hours and days of the total number of respondents will be used for imputation.

This admission meant that for certain observations it was not possible to impute a value that was based simply on the response to C2. For this reason, a three-step method was applied for the imputation of these values:

1. First, a variable for monthly usual hours was calculated according to C2 respondents and to a factor containing the median number of weeks worked by D4 respondents.
2. Then, for those cases with no information for C2, the usual monthly hours were imputed. For this imputation, the conditional median method is followed through the use of the same groups as those used to impute D1.
3. Finally, with the usual hours already imputed, D4 is imputed using the method described in point 2.

With this methodology, the results shown in Table 7 were obtained.

Table 7 Hourly income for both sexes, female and male wage earners, and income gap (2020–2021)

Year	Hourly income Both sexes	Hourly income Women	Hourly income Men	Gender gap
2020	$4,002	$3,797	$4,129	-8.0%
2021	$4,257	$4,403	$4,028	-8.5%

IV.   Calculation of hourly income

Finally, the calculation of hourly income used by the survey corresponds to the ratio between the total income and the total number of hours actually worked during the reference month.

Annex 5. Hypothesis testing, 2021 estimates vs. 2020 estimates

Since the 2020 ESI, a sampling frame different from the sampling frame of previous years (2010–2019) has been used. The variable designating the primary sampling units in the old frame was *id_directorio*, while in the new frame it is *conglomerado* (cluster). In order to make comparisons between years, the 2010–2019 ESI databases have the variable "*conglomerado*", which standardizes the denomination of the primary sampling units in these databases to their new nomenclature in the sampling frame of the 2020 and 2021 ESI. Thus, the databases (2020 ESI or later + the ESI from a previous year) can be merged and the complex design can be stated in statistical software using this variable of clusters homologated to their nomenclature in the 2020 frame[60]. In this way, the measurement of covariance for hypothesis testing can be improved.

In order to evaluate the hypothesis test between estimates from the 2020 ESI or later versus the ESI of a previous year, both databases must be combined through the use of the variable "conglomerado" (cluster) as the variable referring to the primary sampling units.

Hypothesis testing between years; example in R

```r
library(survey)
library(tidyverse)

data_prueba <- bind_rows
(
   esi_2019 %>% select(conglomerado, estrato, fact_cal_esi, sexo,
ocup_ref, ing_t_p, ano_encuesta, factor = factor1, tipo = 1),
   esi_2020 %>% select(conglomerado, estrato, fact_cal_esi, sexo,
ocup_ref, ing_t_p, ano_encuesta, factor = factor2, tipo = 2),
 )

design <- svydesign(id = ~conglomerado, strata = ~estrato, weights =
~fact_cal_esi, nest = TRUE, data = data_prueba)
 options(survey.lonely.psu="remove")

estimacion <- svyby(~ing_t_p, by = ~ano_encuesta, na.rm.all = FALSE,
                    subset(design, ocup_ref == 1 & sexo == 2), svymean,
covmat = T, vartype = "se")

contraste <- svycontrast(estimacion, quote(`2019` - `2020`))

contraste %>%
  as_tibble() %>%
```

---

[60] Because the new sampling frame is dual, this homologation is not complete. Thus, the only clusters that can be homologated are those from the old frame that are also present in the new sampling frame.

```
    mutate("p.value" = pt(abs(c(contraste[1]/SE(contraste)))), df = 60,
lower.tail = FALSE)*2)
```