



NEW NATIONAL EMPLOYMENT SURVEY

**METHODOLOGICAL
AND CONCEPTUAL MANUAL**

SAMPLING DESIGN

April 2010

CHAPTER I: CONCEPTUAL MANUAL

1. Introduction

This document aims to introduce users to the concepts that define the New National Employment Survey (NENE), which replaces, as of April 2010, the former National Employment Survey (ENE) in force since 1966.

The main definitions described in this manual refer to the variables that make up the Labour Legislation Summary (Código Sumario de Empleo, CSE). The definitions take into account updated classification criteria emanating from the resolutions of the International Conference of Labour Statisticians (ICLS), of the International Labour Organization (ILO), including the Twelfth ICLS held in 1982.

As a result of introducing international measurement standards, the conceptual and methodological design of the New National Employment Survey includes, among other aspects, updated concepts of employed and unemployed persons, as well as a distinction between "traditional" and "non-traditional" employed persons.

The updated concepts go together with details of the main ratio indicators, calculated from work activity categories and a summary of the main analytical indicators, consisting of a set of quotients that facilitate analysis and that are linked to the employed population.

2. Classification of working-age population

In the New National Employment Survey (NENE), the classification of the working-age population (WAP) by economic activity or status with reference to the labour force is summarised in the Labour Legislation Summary (Código Sumario de Empleo, CSE).

The overall goal of the CSE is to classify all persons of 15 years old and over according to their employment status, by applying a set of priority rules which, on an international level, are at the base of the approach for measuring employment and unemployment by means of household surveys. These priority rules are translated into a series of questions and sequences that make up the new survey questionnaire.

The conceptual foundations of the CSE are based on updating of the classification criteria according to the resolutions of ILO's International Conferences of Labour Statisticians (ICLS), particularly the Twelfth ICLS held in 1982.¹

As a result of including international standards of measurement, the conceptual and methodological design of the New National Employment Survey results in:

- Updating the concept of employed persons, incorporating the criteria of having worked at least one hour in the reference week and receiving a payment for the work performed.
- Discriminating between "traditional" and "non-traditional" employed, depending on whether the respondent immediately recognizes or identifies his/her activity as work.
- Updating the concept of "unemployed" persons, incorporating the criteria of having sought work within the past four weeks and being available for work in, at least, the two weeks following the interview.
- Classifying the population outside the labour force by reasons for inactivity and according to their potential to participate in the labour market. This is a major difference with the previous survey in which inactivity, based on a self-classification, was associated with the frequency² of the condition. The new instrument classifies the inactive population through a set of questions that are in line with the rule of priority in the classification, which begins with the employed condition and then moves to unemployed. To be economically inactive, a person must not have met the classification criteria of these conditions.

2.1 Concepts and measurement criteria

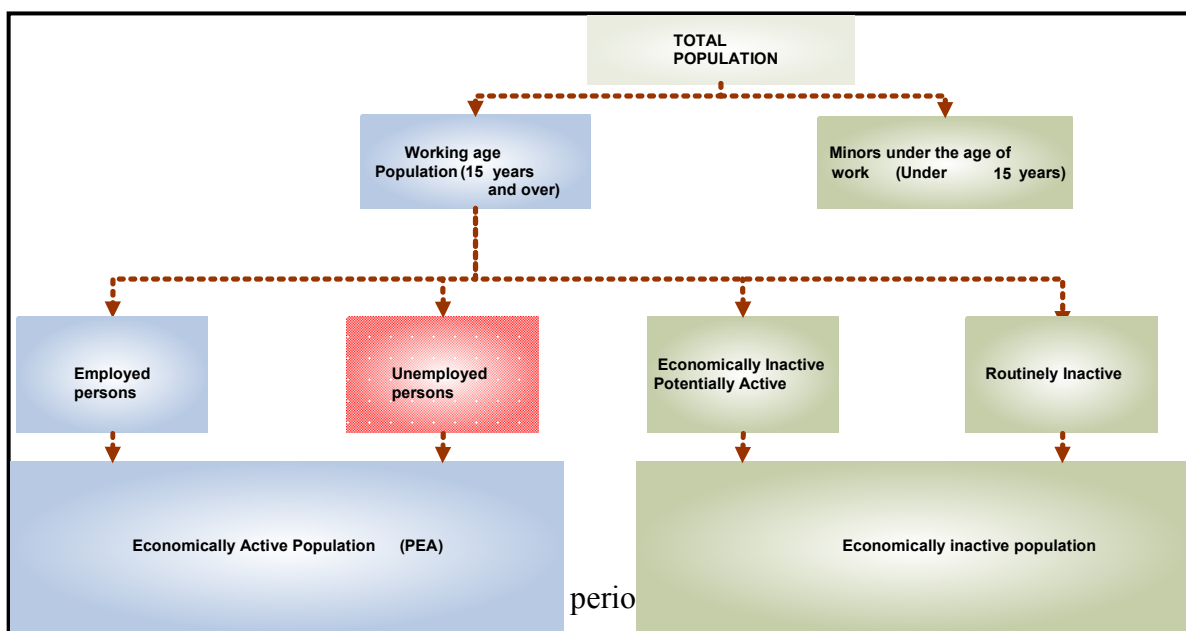
- The following conceptual diagram shows the broad categories into which the population is classified for the purpose of analysing their employment status. The New National Employment Survey is addressed to all persons aged 15 years or more, who constitute the working age population in Chile. People who respond to the questionnaire may be classified as employed, unemployed or economically inactive,³ which are mutually exclusive categories.

¹ "Survey of the Economically Active Population. Employment, unemployment and underemployment." Vol 1. In particular: Part 1. Ch. 2 (points 2, 3 and 4), Ch. 3 (items 2, 3 and 4) and Ch. 6 (item 3).

² The activity status of those over 15 years in the old survey stemmed from what people stated they had done "during most of the previous week" (reference week).

³ Annex 1 presents the sequences from which are constructed each and every one of the Summary Employment Code stocks of the NENE.

Diagram 1:
**CLASSIFICATION OF THE ECONOMICALLY ACTIVE AND INACTIVE POPULATION IN THE
 NEW NATIONAL EMPLOYMENT SURVEY (NENE)**

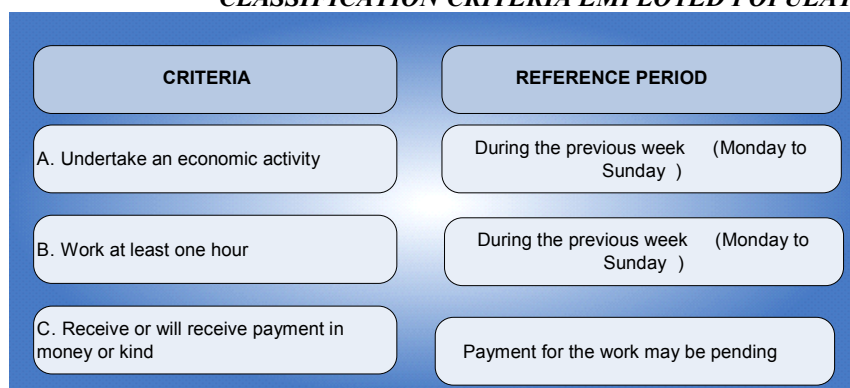


2.1.1 Employed

To qualify as employed, people must meet the following criteria:

- Must have worked one hour or more during the reference week.
- Must have earned (received) a wage, either in pay or profit, for the work undertaken.

Diagram 2
CLASSIFICATION CRITERIA EMPLOYED POPULATION



Work is any activity that is within the production boundary of the System of National Accounts (SNA) and covers all market production and certain types of non-market production (primary products or processed for self-consumption).

i. Traditional and non-traditional employed persons

The difference between "traditional" and "non-traditional" employed persons is subject to the understanding that each person has of the activity performed. When posing questions in section A of the questionnaire, the persons who perceive their work as such will be classified as "traditional employed," according to the above-specified criteria. Meanwhile, those respondents who from the beginning do not identify their activity as work⁴ will be classified as "non-traditional employed" once further inquiries regarding the activity undertaken have been made.

The greater precision in the classification of employed persons allows people to be included who perform work that, by its very nature, whether hours worked, the frequency with which the work is undertaken, or the place of work, would not have been identified as an economic activity in the first place. This precision is particularly relevant in developing countries where, because of cultural reasons, certain activities are not declared even though in many cases they are part of the survival and economic sustainability of households, whether in the short term, during a certain season or form part of the economic cycle.

⁴ Such activity must be paid (except for unpaid family members).

i.i Employed persons absent from work

This definition includes persons who have a job, but that did not work during the reference week for various reasons, although they maintain a formal relationship with their employment and are expected to come back to work after their period of absence.

The international definition specifies certain principles for determining temporary absence from work ("employed persons absent from work"), depending on whether these are wage earning or independent workers, and according to three criteria in line with ILO standards.

- Employed absent from work who maintain a formal job attachment: Persons in this category were absent from work during the reference week for one of the following reasons: holiday or leave; medical leave; part-time or flexible hours; strike; labour dispute; training leave.
- Employed, absent from work, returning soon: Persons who state that they will return to their work within four weeks.
- Employed, absent from work, with wage or profit: Persons who declare that during their absence they will continue to receive income for their work.

The relevance of this classification reflects the fact that absence from work follows different temporary patterns. Depending on the economic cycle, technical or economic reasons can determine the length of absence and return to work. New forms of contract or the timing of some activities may be associated with periods of absence, which are worth identifying for analytical purposes. Temporary absence also exhibits a seasonal bias. During the winter, for example, medical leave might increase, while in summer "vacation or holiday leave" is more frequent.

Question A6 of the questionnaire specifies the reasons of absence from work of employed people during the reference week:

A6 ¿Por qué razón no trabajó la semana pasada?

▼ No se leen las opciones

<input type="checkbox"/> 1. Vacaciones o permisos	
<input type="checkbox"/> 2. Licencia médica	
<input type="checkbox"/> 3. Por horario o jornada variable o flexible	Pasa a B1
<input type="checkbox"/> 4. Huelga, conflicto laboral	
<input type="checkbox"/> 5. Asistencia a cursos de capacitación	
<input type="checkbox"/> 6. Problemas de salud	
<input type="checkbox"/> 7. Suspensión temporal del trabajo	
<input type="checkbox"/> 8. No tuvo pedidos (clientes)	
<input type="checkbox"/> 9. Razones climáticas o catástrofes naturales	
<input type="checkbox"/> 10. Su trabajo es ocasional	
<input type="checkbox"/> 11. Razones económicas o técnicas de la empresa o negocio (rotura de equipos, falta de materias primas o capital para comprar)	
<input type="checkbox"/> 12. Su trabajo es estacional	
<input type="checkbox"/> 13. Clausura de negocio o de empresa	
<input type="checkbox"/> 14. Otras razones (especificar) _____	

A6 Why did you not work in the previous week?

1. Holiday or authorised leave
2. Sick leave
3. Variable or flexible working hours
4. Strike, labour conflict
5. Taking part in training course
6. Health problems
7. Temporarily suspended from work
8. No new orders (customers)
9. Due to the weather or natural disaster
10. The work is part-time
11. Economic or technical reasons affecting the firm or business (mechanical breakdown, shortage of raw materials or the capital to purchase these)
12. Work is seasonal
13. Business or the company was closed down
14. Other reasons (specify)

2.1.2. Unemployed persons

To qualify as unemployed, people must meet the following criteria:

- To not have worked during the reference week.
- To have sought a job in the last four weeks (including the reference week).
- Be available to work in the next two weeks following the reference week.

Diagram 3
CLASSIFICATION CRITERIA UNEMPLOYED POPULATION

CRITERIA	REFERENCE PERIOD
A. Not have worked 1 hour	During the previous week (Monday to Sunday)
B. Actively searched for work	During the last four weeks
C. Is available to work	From end of reference week and up to 15 days afterwards

The last two criteria allow us to define the "pressures of supply" in the labour market. Looking for a job, but not being available means that, in fact, persons are not competing for a job post. Being available but not looking for a job means that the person has not approached the labour market as a supplier.

Within the unemployed, we set apart persons without work from first time job seekers, depending on whether the unemployed person has had a former job or not.

2.1.3 Outside of the labour force or inactive

The economically inactive population consists of people who did not have a job and neither sought and/or were available for work. That is, this category includes all persons not classified as employed or unemployed.

i. Inactive by reasons and new subcategories

With the New National Employment Survey, inactive persons are classified according to reasons of inactivity (not searching or unavailable for work, or both simultaneously). This is a conceptual change that refers to people's reasons for not entering the labour market and that allows inactivity to be understood not as a permanent state but as a condition that can change for a number of reasons in the short-time.⁵ These reasons might be linked to the family, study, retirement, pension or survivor's benefit, and permanent health problems. Additionally, the economically inactive are categorised according to reasons of discouragement; seasonal circumstances; temporary personal reasons, or because they are job starters.

⁵ This clarification is associated with the "reference week" being considered as a short period during which the labour situation was surveyed, according to the conceptual framework for the labour force on which employment and unemployment measurements are based.

E9 ¿Por qué razón no buscó un empleo o no ha hecho preparativos para iniciar o reanudar una actividad por cuenta propia durante las últimas cuatro semanas?

▼ No se leen las opciones

- 1. Iniciará pronto una actividad por cuenta propia **→ Pasa a E10a**
- 2. Encontró un empleo que empezará pronto **→ Pasa a E10b**
- 3. Por responsabilidades familiares permanentes (cuidado de hijos o personas dependientes)
- 4. Está estudiando o preparando estudios
- 5. Es jubilado/a
- 6. Es rentista
- 7. Es pensionado/a o montepiada
- 8. Por motivos de salud PERMANENTES
- 9. Espera la estación de mayor actividad
- 10. Por motivos de salud TEMPORALES **Pasan a E11**
- 11. Por responsabilidades familiares de carácter temporal
- 12. Está embarazada
- 13. Espera los resultados de un proceso de selección o que lo llamen
- 14. Algún miembro del hogar no se lo permite
- 15. Cree que por su edad no le darán empleo
- 16. Cree que no lo encontrará
- 17. Se cansó de buscar
- 18. Cree que ningún empleo o actividad se adapte a su calificación
- 19. Le piden demasiados trámites para iniciar una actividad por cuenta propia
- 20. No sabe donde dirigirse ni a quien acudir
- 21. No quiere, no necesita trabajar **→ Pasa a E13**

E9 Why did you not search for work or make any effort to start or resume an own-account activity during the last four weeks?

- 1. Will soon start an own-account activity
- 2. Found a job and will start soon
- 3. Due to permanent family responsibilities (caring for a child or other dependents)
- 4. Studying or preparing to study
- 5. Is retired
- 6. Is a rentier
- 7. Receives a pension or survivor's benefit
- 8. Due to permanent health problems
- 9. Waiting for a season with more employment activity
- 10. Due to temporary health problems

11. Due to temporary family responsibilities
12. Due to pregnancy
13. Waiting for the response (written or phone call) from a selection process
14. Prohibited by a member of the household
15. Believes that due to age will not be offered work
16. Believes that she/he will be unable to find work
17. Became tired of searching for work
18. Believes that no job or activity suits their qualifications
19. Too much paperwork involved in starting up an own-account activity
20. Does not know where to go or who to ask
21. Does not want or need a job

ii. Inactive persons who are potentially active

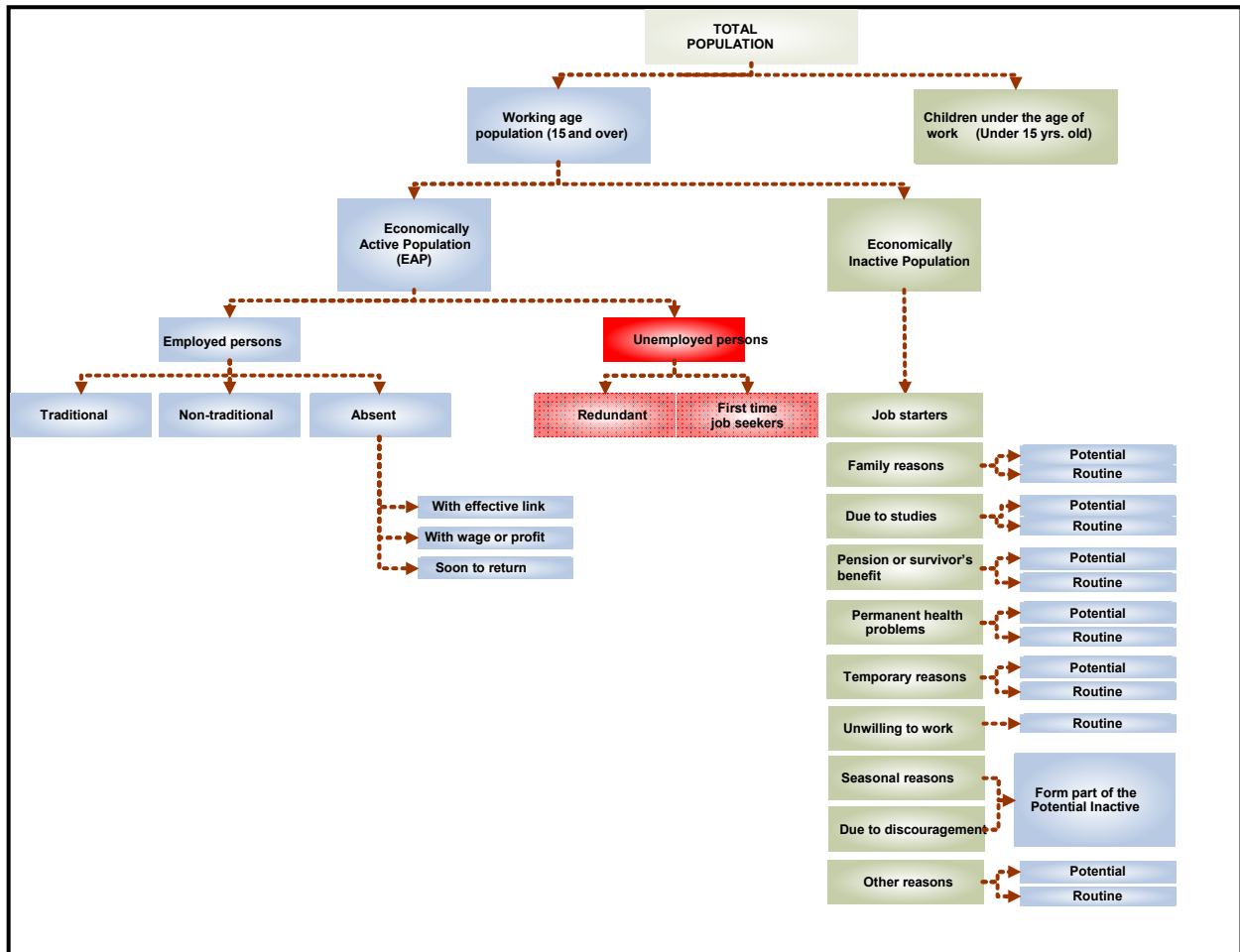
The economically inactive who are potentially active are those persons who, regardless of the reason they are considered inactive, have searched for a job, but would not be available to work or have not searched, but would be available if a job was offered. The potentiality of economically inactive persons is assessed by their confirmation or disposition, respectively.

The word "potential" refers to a presumption of the possibility that a person who is currently not putting pressure on the labour market may in fact do so at some point.

Those inactive persons that have neither sought a job nor would be available to work are identified as the routinely inactive.

It is worth noting the new categories of economically inactive due to "discouragement" and "seasonal" factors, which are terms that have become increasingly important at international level when analyzing the behaviour of the labour force; these inactive persons correspond to population groups that, due to their characteristics, have a stronger link with the labour force compared to other inactive groups. At international level, this concept is referred to as being "marginally attached to the labour force." According to the New National Employment Survey, these categories correspond to those people classed as "economically inactive but potentially active."

Diagram 4
CLASSIFICATION OF ECONOMICALLY ACTIVE AND INACTIVE POPULATION ACCORDING TO
THE NEW NATIONAL EMPLOYMENT SURVEY (NENE)



ii.1 Job Starters

These are persons who have not sought employment in the previous four weeks as they will “soon begin” a job, and who may or may not be available to work in the two weeks following the reference period.⁶

If available for work, OECD countries include such people in the category of unemployed persons. If not available, they are considered as economically inactive as they have neither sought a job or are available to work. In general, availability implies greater willingness of participation in the labour force in the short-term, although, in essence, being a “job starter” (available or unavailable for work) is an intermediate stage between inactivity and joining

⁶ Also known as people with “future incorporation” in the labour force. ILO, 1990

the labour force.⁷ According to the classification of inactive persons in relation to their potential to become part of the labour force, “job starters” is the subgroup that has most potential.

ii. 2 Due to seasonal factors

The economically inactive due to "seasonal factors" are those persons who state that they are waiting for "the season of most employment activity," and who at the same time indicate they are available for work.

Their level of potential to join the labour force is high, and in some countries this subgroup is included in the "marginally attached" population, consisting of the group that is most likely to enter the labour force in the short-term or as soon as labour market conditions improve.

Although this is not a large group, it can have an impact on the indicators measuring pressure on the labour force.

ii. 3 Due to reasons of discouragement

Discouraged workers are persons who did not seek employment in the last four weeks, who desisted because they thought they would not find a job, but who would be available to start work within the two weeks following the reference week. Like those who are inactive due to seasonal factors, in some countries this group is included in the “marginally attached” bracket, also representing its largest group.

ii.4 Other reasons

This is a complex category of economically inactive people in terms of composition, as it comprises all those individuals who do not fully qualify in another category of inactivity. However, unlike the groups mentioned below, potentiality to enter the labour force predominates in this segment, either because they have sought employment or because they claim to be available to work. In fact, some persons that claim they will undertake a job/own-account activity in the short term are not considered "job starters" because, at the same time, they state that they have not taken specific measures nor made arrangements to take up a job, which would confirm their condition as “job starters.”

⁷ Job starters who are available for work, according to special considerations of the ILO and OECD adaptations, can be considered as unemployed persons. There is no standard guideline regarding this point, therefore the New National Employment Survey will consider them as a group belonging to the economically inactive population, and as unemployed (standardised unemployment rate) for the purposes of comparability with other OECD countries.

ii. 5 Due to temporary personal reasons

The economically inactive grouped in this category state that they have not sought work due to health issues, pregnancy or temporary family responsibilities. If they declare that they are available for work they are classified as "inactive who are potentially active."

ii. 6 Due to permanent family reasons

This category covers a significant segment of the economically inactive population, predominated by a routine condition⁸ of inactivity, and mostly composed of women.

The persons in this subgroup are "typically" inactive in terms of presenting a low predisposition to join the labour market. However, if they state that they are available for work, they are then classified as "inactive potentially active."

ii. 7 Due to reasons of study

This is another significant category of economically inactive persons, associated with the status of being a student. Those within this group who state that they are "available to work" represent a small proportion of the total, thus this segment has a low potential of entering the labour force.

ii. 8 Pension or Survivor's benefit

Persons in this category are those who receive a permanent income as a result of the work performed during their active working life, or because of a pension received by a surviving spouse. The potential of this group to join the labour force is very low, with a predominately routine condition.

ii. 9 Due to retirement

This is a group that shares similar characteristics to those receiving a pension or survivor's benefit. Basically, these are persons who receive a permanent income derived from their working life.

ii.10 Permanent health problems

People in this group indicate that they either did not seek employment, or that they were unavailable for work because of health problems preventing them from doing so. This category of inactive persons is basically a group that has virtually no potential to enter the labour force, and only a minimum of such people actually become available for work.

⁸ Have not sought employment during the previous four weeks and are unavailable to work in the two weeks following the reference period.

ii. 11 Unwilling to work

These are persons who state that they do not wish to work; consequently their potential to join the labour market is zero. In general, these are individuals who do not need to work as they have a guaranteed regular income.

APPENDIX 1: QUESTIONS FLOW CHART NEW NATIONAL EMPLOYMENT SURVEY (NENE)

This Appendix presents the flow charts that illustrate the sequence of questions from which are derived the different categories of the **Employment Summary Code (Código Sumario de Empleo - CSE)**. The result is a nominal and exclusive classification of the working age population (see Diagram 4 in the main text).

QUESTIONS FLOW CHART NEW NATIONAL EMPLOYMENT SURVEY QUESTIONNAIRE

The Specific Employment Summary Code (**Código Sumario de Empleo Específico - CSE_E**) is the most disaggregated version of the CSE and is established on the basis of modules A and E of the New National Employment Survey questionnaire (NENE). It has a complex structure, and thus requires that both modules are responded to at the same time.

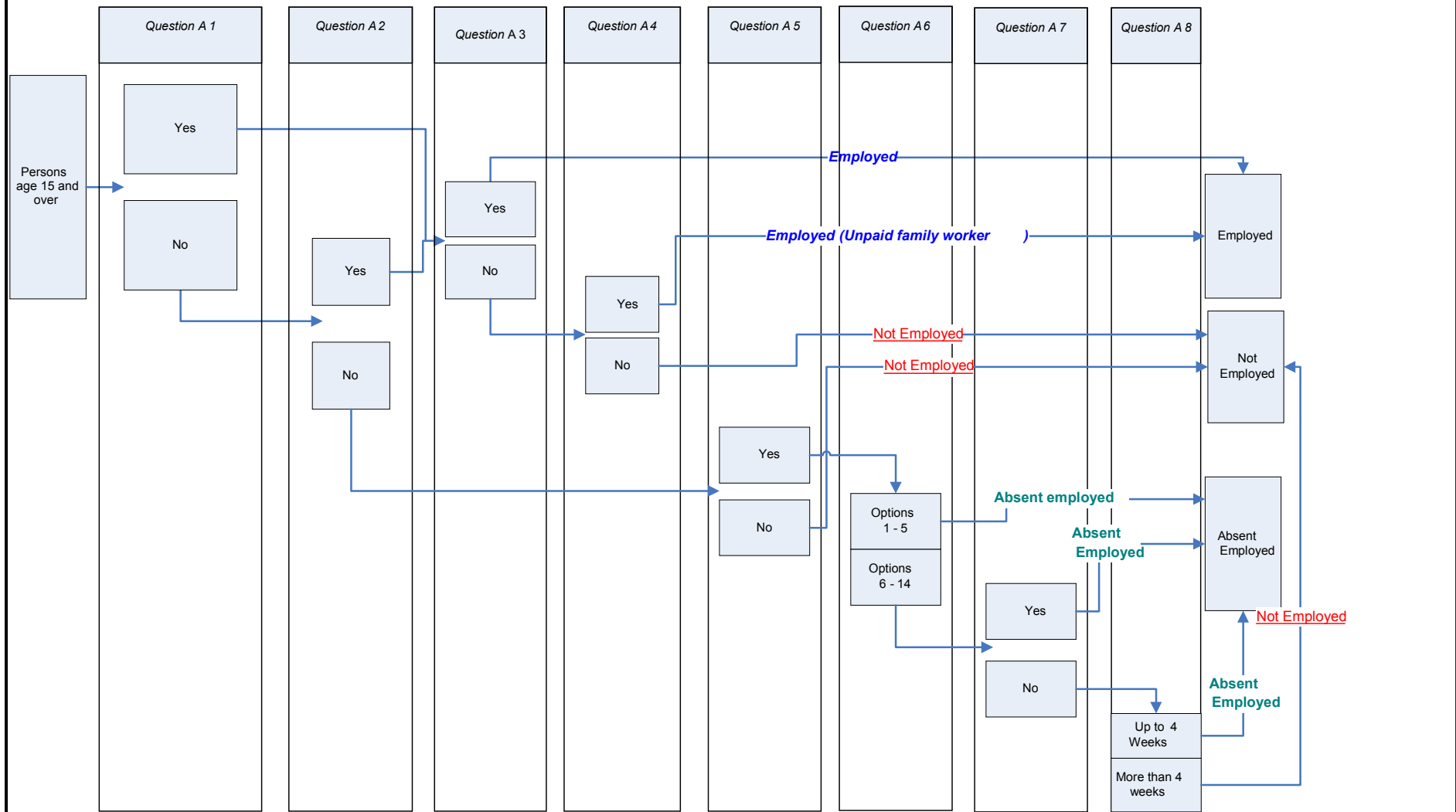
Module A establishes the stock of employed persons among the population aged 15 and over in the country.

“Non-employed” persons must pass to Module E in order to determine to which of the different stocks of the population aged 15 and over they belong to.

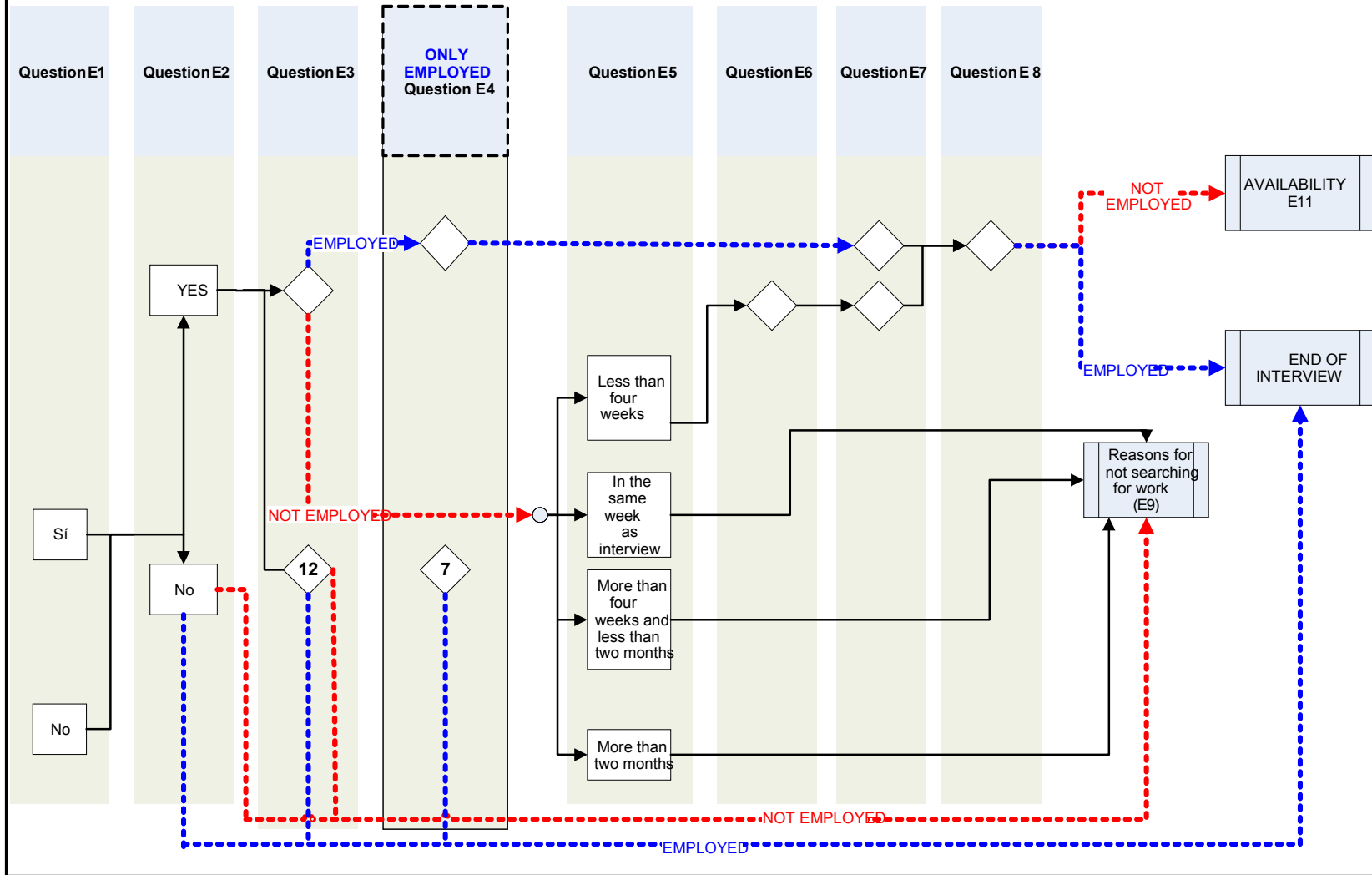
The sequence of questions for Module A, followed by those for Module E, are as follows:

Module A: LABOUR SITUATION IN REFERENCE WEEK

Questions A1 to A8

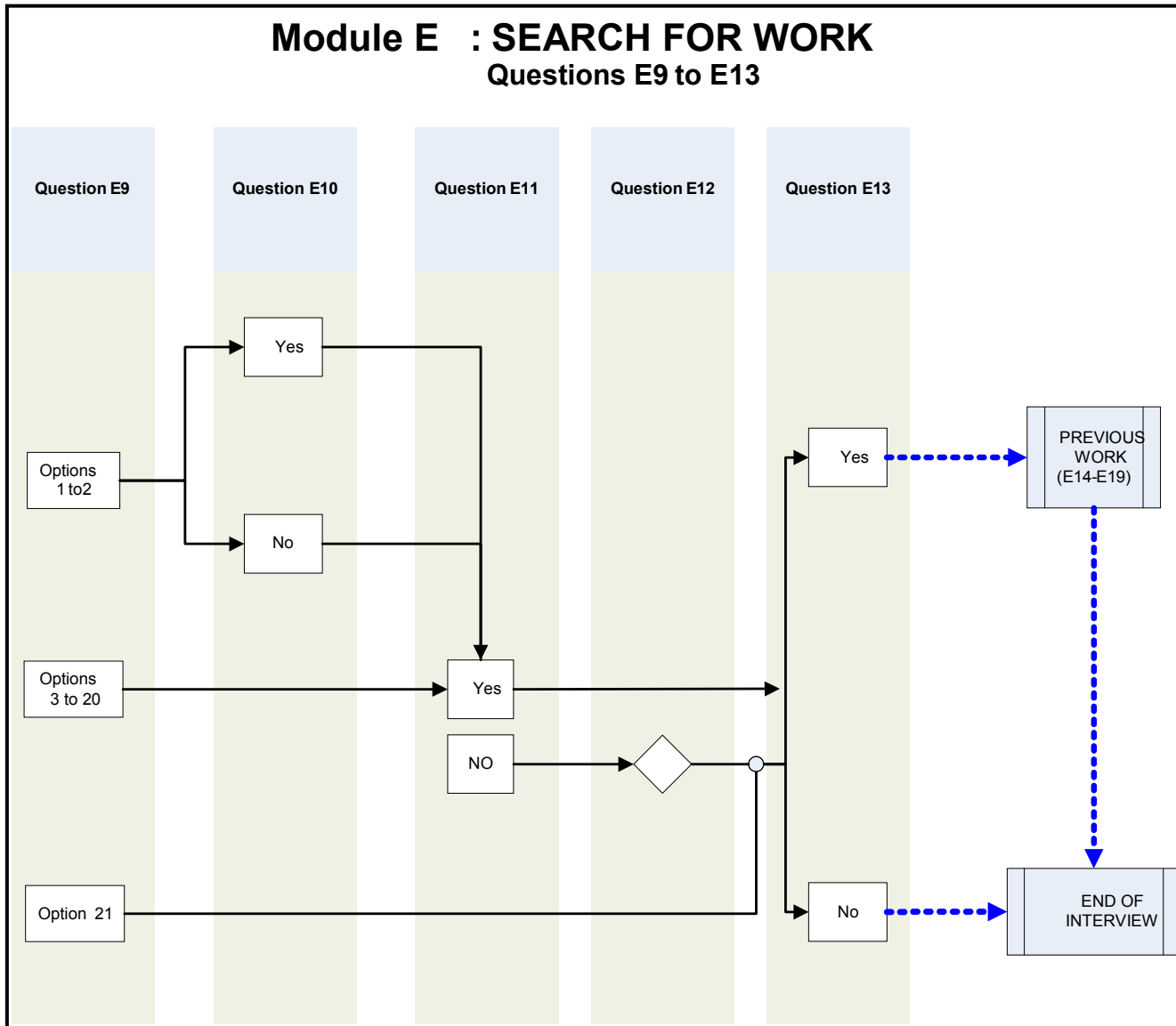


Module E : SEARCH FOR WORK Questions E 1 to E 8



Module E : SEARCH FOR WORK

Questions E9 to E13



APPENDIX 2: GLOSSARY

Working-age population (WAP): population aged 15 and over currently residing in the country.

Economically-active population (EAP) or labour force: people of working age who, during the reference week, are eligible to be included in the categories of employed or unemployed.

Employed: all persons of working age who during the reference week worked for one hour or more, for pay or profit, in the context of an employee/employer relationship or own-account worker.

- I. **Traditional employed:** employed persons who from the beginning of the survey responded affirmatively to the question of whether he or she had worked the previous week.
- II. **Non-traditional employed:** employed persons who first said that they did not work the previous week, but then, when asked more inclusive and open questions, state that they did work.
- III. **Absent Employed:** employed persons who during the reference week did not attend work for various reasons, although they remained closely associated with their job.

Unemployed: All persons in the working-age population who during the reference week were without a paid job, available for work and had either actively sought work in the past four weeks ending with the reference week, or had a new job to start within the next four weeks.

- I. **Without work or jobless:** any person who, having met the criteria for being unemployed, previously had a job that lasted one month or more.
- II. **First-time job seeker:** any person who, having met the criteria for being unemployed, have not had a job that lasted at least one month.

Available job starters: all persons in the working-age population who state that they have not sought work in the previous four weeks because they are about to start a job and, at the same time, declare they are available for work.

This group is included in the economically inactive population; however, for the purpose of statistical harmonization with the OECD, it is considered as part of the labour force.

Economically inactive population: All persons of working age, who are neither employed nor unemployed. This category comprises people that declare any of the following reasons for inactivity:

- Job starters⁹
- Seasonal reasons
- Reasons of discouragement
- Temporary reasons
- Permanent family reasons
- Study
- Pension or survivor's benefit
- Retirement
- Permanent health problems
- No desire to work

⁹ Persons who reached an employment agreement or made arrangements to start an own-account activity in the previous four weeks, and who may or may not be available for work. If they are available, they are considered part of the economically active population for the purpose of OECD standardised rates.

APPENDIX 3: MAIN RATIO INDICATORS

The main ratio indicators are constructed from the categories of activity status:

Unemployment rate: number of unemployed persons expressed as a percentage of the labour force.

Standardised unemployment rate: Number of unemployed persons plus available job starters, expressed as a percentage of the labour force (including available job starters). That is, it is added in both the numerator and the denominator to available job starters.

Labour force participation rate: number of persons in the labour force as a percentage of the working-age population.

Employment rate: number of employed persons as a percentage of the working-age population.

General labour force pressure rate:

Number of employed persons seeking work, plus unemployed persons, plus available job starters, expressed as a percentage of the labour force (including available job starters). That is, employed persons that have looked for a job are included in the numerator, and available job starters are included in both the numerator and the denominator.

APPENDIX 4: ANALYTICAL RATIO INDICATORS

This is a set of quotients that facilitate the analysis of the employment situation and that are associated with the employed population. The purpose of these quotients is to improve the characterisation of conditions of employment of the population already employed in terms of: hours worked and labour benefits.

Percentage of involuntary part-time employed persons: Number of people working involuntarily¹⁰ in a part-time capacity,¹¹ expressed as a percentage of the total number of part-time employed persons.

Percentage of voluntary part-time employed persons: Number of people voluntarily working part-time¹², expressed as a percentage of the total number of part-time employed people. In other words, this is an adjunct to the preceding indicator.

Percentage of employed persons working more than 45 hours per week: Number of persons normally working more than 45 hours per week, expressed as a percentage of the total number of employed persons.

Percentage of wage-earning workers with a written contract: the number of wage-earning workers who have a written contract, expressed as a percentage of total Wage earners.

Percentage of wage-earning workers with a fixed-term contract: the number of wage-earning workers who have a fixed-term contract, expressed as a percentage of total Wage earners.

Percentage of wage-earning workers with a high level of labour benefits: the number of wage-earning workers with a written contract, pension contributions, health insurance and unemployment insurance, expressed as a percentage of total Wage earners.

Percentage of wage-earning workers with a low level of benefits: the number of wage-earning workers that do not have a written contract, or make pension contributions, and have no health and unemployment insurance, expressed as a percentage of total wage earners.

¹⁰ As they state that they would like and are available to work more hours per week.

¹¹ Normally no more than 30 hours per week.

¹² Normally no more than 30 hours per week.

APPENDIX 5: OTHER DEFINITIONS

Normal hours: the hours normally worked in a determined reference period.

Actual working hours: actually worked hours quantify the time spent by workers in activities that contribute to the production of goods and services, during a determined reference period.

Involuntary part-time work: Involuntary part-time workers are those employed people who state that they have a part-time job, but who wish to work more hours and are available to do so (either immediately or within two weeks following the reference period).

APPENDIX 6: CLASSIFICATIONS USED IN THE SURVEY¹³

Classification by employment status

The International Classification of Status in Employment (ICSE), adopted by the ILO in the Thirteenth ICLS in 1982, is used to classify the economically active population. This system classifies said population in categories such as Employer; Own-account worker; Employee who receives a wage or salary, and Unpaid family worker.¹⁴

Industrial Classification of Economic Activities

The NENE uses the International Standard Industrial Classification of all Economic Activities (ISIC Rev.3). This is a standardized classification of economic activities, aimed at classifying entities according to their main economic activity. This classifier was approved by the UN Statistical Commission 1998, which set the goal of implementing a classification system that fostered international comparability.

The general structure of ISIC Rev.3 consists of 17 classifications represented by letters ranging from A to Q; 60 divisions, 159 groups and 292 classes. The first two digits indicate the division, three digits identify the group and four digits taken together indicate the class. For classification purposes, estimates provided by NENE represent the 17 levels of classification.

International classification of occupations

The International Standard Classification of Occupations (ISCO-88), published by the ILO in 1988 and replacing ISCO-68, is used in reference to the classification of occupations. This is a system used to classify data and information on occupations, designed as a national classification tool and to facilitate international comparability.

The general structure of ISCO-88 is composed of 10 major groups, 28 sub-major groups, 116 minor groups and 390 unit groups. Disaggregation to one digit indicates the major group; to 2 digits the main sub-major groups; to 3 digits the minor groups, and to 4 digits the unit groups. For classification purposes, estimates provided by NENE are at the level of the major groups.

¹³ For more detailed information on the first 3 classifiers, see the nomenclature assisted search engine (BAN) at the following link: http://www.ine.cl/pop_up/nomenclaturas.php

¹⁴ See the Conceptual Manual for more detail.

Education classification

To classify the population according to their educational level the International Standard Classification of Education (ISCED). This instrument was designed by UNESCO in the early 1970s for compiling and presenting educational statistics in individual countries and for international comparability. This classifier basically covers two classification variables: levels and areas of education.

* For more detailed information on these 3 classifiers, see the Nomenclature Assisted Search Engine (BAN). http://www.ine.cl/pop_up/nomenclaturas.php

CHAPTER 2: SAMPLING DESIGN MANUAL

1 Introduction

Within the context of continuous improvements of the statistics produced by the National Statistical Institute (INE) and the interest in adopting the best practices used by OECD countries, the Department of Research and Development of INE has created a Sampling Design in order to innovate the measurement of employment, in order to make estimates comparable with other member countries of the OECD.

This new way of measuring employment and its characteristics uses two sampling frames by area (Urban, RUA¹⁵ and Rural). The sampling frame used in urban areas is the Master Sampling Frame, which is continually updated for this area. The sampling frame used in RUA and rural areas is that traditionally used by the former National Employment Survey - a frame that refers to the Census, in this case the 2002 Census, with continuous updating of sections belonging to the sample.

The sample sizes used in this new survey are calculated so as to provide within the levels of estimation similar errors to those reported in the previous employment survey, considering these as acceptable in relation to what has been normally observed over time.

¹⁵ RUA: Rest of Urban Area. Urban settlements with less than 40,000 inhabitants.

2 Goal of the survey

The New National Employment Survey (NENE) aims to characterise and quantify the population aged 15 and over regarding their status in the employment market.

This characterization allows information to be gathered in a similar way to surveys undertaken by other OECD countries, including the transformations and developments in the labour force, specifically the Unemployment rate, both at domestic level and in the corresponding Levels of estimation.¹⁶

3 Target population

The target population is made up of all persons in the country living in occupied private dwellings. This definition excludes all people living in collective dwellings such as hospitals, prisons, convents, military quarters and others, but includes people residing in private dwellings inside such facilities, such as concierges, caretakers and others. Furthermore, people living in Areas of Difficult Access (ADAs) area excluded from the geographical coverage.

¹⁶ For more information on levels of estimation see Table 3.

4 Sampling Frame and stratification

4.1 Sampling frame: Conformation, segmentation and updating of the Sampling Frame

In urban areas, the New National Employment Survey uses an updated sampling frame of residential blocks. This is created using high resolution digital maps taken from related services such as Mapcity and Dmapas, and on data provided by the different municipalities regarding building permits, in addition to cartographic material obtained from the 2001 Pre-census survey and the 2002 Population and Housing Census (regional, provincial, municipal and district maps).

As for RUA and rural areas, the sampling frame is created based on information provided by the above described 2002 Population and Housing Census, with continuous updating of the sections belonging to the sample.

The sampling frame used for surveying in urban areas incorporates a new concept of sampling unit, the residential block, the configuration of which corresponds to natural geographic boundaries, unlike the sampling frame used for RUA and rural areas, which is composed of sections established according to sets of a given number of dwellings (see Table 1); of the stability presented by these sections over time (according to data from the 2002 Census); and easy identification in the field (respecting communal boundaries).

Table 1 - Section size and number of dwellings to be surveyed by RUA and rural areas

Area	Number of dwellings	
	By section	To be surveyed by section
Rest of Urban Area (RUA)	50 - 150	12
Rural	40 - 120	15

4.2 Stratification of the Sampling Frame

In both cases, the sampling frame is stratified according to the geographical condition (Administrative-Political Division) and by the number of dwellings and population established by the 2002 Census, as well as segregation depending on the economic activity prevailing in the area.

This information allows strata to be classified according to:

- Cities or large Urban Centres (UC): Made up of cities or groups of adjacent cities with 40,000 inhabitants and over.
- Rest of Urban Area (RUA): Composed of Urban Centres with less than 40,000 inhabitants.

- Rural Area (R): Made up of the cluster of entities classified as rural according to their population size -less than 1,000 inhabitants, or between 1,001 and 2,000 inhabitants with a predominance of Economically Active Population (EAP) engaged in primary activities.¹⁷

The sampling frame of residential blocks respects the geographical stratification and segregation based on the economic activity of sampling frames previously used by the INE,¹⁸ and also incorporates a new concept of stratification to obtain samples of residential blocks of different sizes. Thus, in the construction of the sampling frame we have included a variable that classifies residential blocks according to their size in terms of number of dwellings, forming 5 groups denominated Strata of Size, and 30 subgroups. This allows residential blocks of different sizes to be proportionally represented in the sample, selecting a fixed number of dwellings as appropriate to the group and subgroup (see Table 2).

Table 2 - Classification of residential blocks in strata of size according to the number of dwellings in the sample and the number of dwellings to be surveyed, determined by the sample size of the subgroup

Group	Subgroup	No. of dwellings by residential block		Dwellings to be surveyed
		Minimum	Maximum	
1	1	8	9	2
	2	10	13	3
	3	14	17	4
	4	18	23	5
2	5	24	25	6
	6	26	29	7
	7	30	33	8
	8	34	37	9
	9	38	44	10
3	10	45	45	11
	11	46	49	12
	12	50	53	13
	13	54	57	14
	14	58	61	15
	15	62	65	16
	16	66	69	17
	17	70	73	18
	18	74	77	19
	19	78	81	20

¹⁷ By primary activity we consider all activities associated with the exploitation of natural resources (agriculture, hunting, fishing, mining, etc.).

¹⁸ For more information see the Methodology of the 2006 National Employment Survey on the INE website.

4	20	82	85	21
	21	86	89	22
	22	90	93	23
	23	94	97	24
	24	98	102	25
	25	103	105	26
	26	106	109	27
	27	110	113	28
	28	114	154	29
5	29	155	303	30
	30	304	1043	31

* A sampling fraction of approximately 25% is considered.

5 Description of sample design

The preparation of the sample design includes the study of the variables to be analysed, in this case, the behaviour of the labour force, and specifically the Unemployment rate.

The sample design used is defined as probabilistic, two-stage and stratified. Accordingly, the selection probabilities differ depending on whether the area is urban, other urban (RUA) or rural:

Urban Area: The selection probability of first-stage units (residential blocks) is proportional to its size in terms of number of second-stage units (dwellings) in the stratum (geographic and size) and the number of dwellings in each unit. The selection of second-stage units (dwellings) is carried out with equal probability and made using a systematic selection methodology, implemented in the field, starting with one randomly selected dwelling, and then using a constant to select the other units.

RUA and rural area: The selection probability of first-stage units (sections) is proportional to its size in terms of the number of second-stage units (dwellings) in each unit. As for the selection of second-stage units (dwellings), it is carried out with equal probability and with a systematic selection methodology previously defined in-house; that is, the data gatherer goes to the field with a specific directory of the dwellings he/she will survey.

6 Geographic coverage and estimate areas

As mentioned previously, the geographical coverage or scope includes the entire population living in Chile with the exception of those in Areas of Difficult Access (ADAs); this means that it covers the bulk of the population living in Chile's continental territory.

The areas of estimation covered by the sample generally correspond to the following:

National level	Total : Urban and rural
	U : Urban ¹⁹
	CD : Large cities
	R : Rural
Regional level	Total : Urban and rural
	U : Urban
	CD : Large cities
	GC : Large urban centres ²⁰
	R : Rural
Provincial level	Total : Urban and rural

7 Sample and information units

The sample units depend on the area concerned. In terms of the selection methodology of the sample, there are two sample units. First-stage sample units for urban areas (cities) correspond to "residential blocks", while for RUA and rural areas these units correspond to "sections." Second-stage sampling units, in turn, correspond in both cases to "occupied private dwellings" within the selected residential blocks or sections.

The units of analysis or information are all those people living in the selected dwellings, although only those aged 15 and over are surveyed, and even when certain characteristics are registered (household registration card) pertaining to each of the persons living in the selected household.

8 Establishing the sample size

Considering the purpose of the survey and in accordance with the levels at which representative and reliable estimations of the variable to be measured need to be obtained, the sample size is determined from an average unemployment rate -based on estimated

¹⁹ The Urban estimate level include large cities and the Rest of Urban Area (RUA)

²⁰ This estimate level corresponds to a group of large cities within a given region, such as the Greater Valparaiso, Greater Santiago and Greater Concepción areas.

rates from 5 moving quarters from a pilot study conducted in 2008- and from obtaining similar sampling errors to those reported by the former employment survey at the same levels of estimation (see Appendix).

As such, the size by level of estimation was calculated using the following formula:

$$n \cdot \bar{m} = \frac{S^2 \cdot Deff}{(Cv/100) \cdot \bar{Y}}$$

Where:

$n \cdot \bar{m}$: Sample size or total number of dwellings in the level of estimation.

n : Number of sections or residential blocks by stratum.

\bar{m} : Number of dwellings to be surveyed by section or residential block.

S^2 : Average population quasi-variance of the number of unemployed persons⁶ obtained from five moving quarters of the pilot study.

$Deff$: Design effect, obtained in the pilot study.

\bar{Y} : Average unemployment rate of five moving quarters.

The sample allocation at different strata considers, besides the pyramid structure²¹ of the sampling errors, other factors pertaining to each stratum:

- a) Unemployment rate within each stratum.
- b) Coefficient of variation of the number of unemployed persons²²

Estimates based on primary units in the new sample frame are more variable, in part due to the heterogeneity of the sizes; a larger sample size than that used in the former employment survey is required to obtain the same sampling errors at the same level of estimation. Due to budgetary constraints, it was decided to decrease the level of confidence from 95% to 90% in order to keep both similar sizes as sampling errors.

9 Sample distribution

The total sample is distributed over a whole quarter and split into three sub-samples approximately similar in size; each one is assigned to one of the three months of the moving quarter. Therefore, three subsamples are created, one for each month, and each one by itself is not accurate enough for all estimate levels.

With this distribution, all dwellings surveyed in month "t", are surveyed again in month "t+3" and each dwelling in the sample is surveyed once per quarter.

As such, estimates for a whole quarter are computed using data from months "t", "t-1", and "t-2".

²¹ This pyramid structure means that sampling errors are higher as the level of estimation is disaggregated.

²² As an approximation of the coefficient of variation of the Unemployment rate.

10 Sample rotation

The rotation of the sample is a procedure aimed at keeping the sample up to date and avoiding overusing respondents.

To perform the rotation, the sample of sections and residential blocks is divided into rotation shifts, which are structured considering 1/6 of the urban sample (residential blocks), 1/9 of the RUA sample and 1/12 of the rural sample (sections). This enables the rotation of the entire sample over an 18-month period in the urban area, 27 months in the RUA area and over a 36-month period in the rural area.

The rotation consists of replacing dwellings within the residential block or section; when no further substitutions are available, the whole residential block or section is replaced by another with similar characteristics.

11 Selection of sampling units and expansion factors

11.1 Selection of units for the first and second stages

The selection of first-stage units, in line with the two-stage method, was performed in each stratum with probability proportional to the size of the unit (sections or residential blocks) in number of dwellings. Thus, the probability of inclusion of the *i*-th first-stage unit is equal to:

$$\pi_{hi} = n_h \cdot \frac{M_{hi}}{M_h}$$

where:

h : Represents the stratum.

i : First-stage units (residential block or section).

n_h : Number of first-stage units in the sample, in stratum *h*.

M_{hi} : Number of dwellings in the first stage unit “*i*”, within stratum *h*, according to the 2002 Census.

M_h : Number of dwellings according to the 2002 Census (for RUA and rural areas) or when taking the sample from the updated urban frame (for the urban area (CD)).

The selection of second-stage units (dwellings within each one of the selected first-stage units) was undertaken with equal probability for all dwellings in the corresponding unit, using a systematic selection. This probability is:

$$\pi_{j/hi} = \frac{m_{hi}}{M'_{hi}}$$

where:

j/hi : Represents dwelling “*j*”, since first-stage unit “*i*” was selected within stratum “*h*”.

m_{hi} : Number of dwellings to be chosen in each first-stage unit.

M'_{hi} : Updated number of dwellings in the first-stage unit “i” within stratum “h”.

11.2 Factors of expansion

The factor of expansion refers to the number of dwellings in the population represented by a single dwelling within the sample. This is informed to the persons living in the dwelling.

According to the design, this factor corresponds to that applied to a two-stage sample. From a sampling perspective this depends on the number of dwellings included in first-stage units (residential blocks or sections) and the number of these in each stratum; and in terms of population, it depends on the number of dwellings in the stratum and in the selected section or residential block.

The theoretic factor of expansion $F1_{hi}$ is calculated as the inverse of the probability of selection of each dwelling. This factor is constant by section or residential block and is applied to all dwellings and therefore to all persons who belong to the section or residential block i of stratum h.

The calculation is as follows:
$$F1_{hi} = \frac{M_h}{M_{hi} \cdot n_h} \cdot \frac{M'_{hi}}{m_{hi}}$$

where:

M_h : Number of dwellings in stratum h by the 2002 Census (in the case of sections) or by the year the sample frame has been updated in the case of residential blocks.

M'_{hi} : Updated number of dwellings in the first stage unit i within stratum h.

M_{hi} : Number of dwellings in the first-stage unit i within stratum h by the 2002 census or by the year the urban frame is updated.

n_h : Number of first-stage units in the sample in stratum h.

m_{hi} : Number of dwellings in the sample in the first-stage unit i within stratum h.

The factor of expansion includes an adjustment of population by stratum, where each stratum has its own estimates of population projections by sex and age bracket (under 15 years and 15 years and older) in order to increase the precision of the estimates and be consistent with population stocks of population projections. Thus, after this adjustment, for each section or residential block in stratum h, there are 4 factors of expansion that apply to all persons in the section providing they are: men under 15 years old; men aged 14 or 15 and over; women under 15; and women aged 14 or 15 and over.

The adjusted factor is calculated by the following formula:

$$F2_{hi} = \frac{M_h}{M_{hi} \cdot n_h} \cdot \frac{M'_{hi}}{m_{hi}} \cdot \frac{P_h}{\hat{P}_h} = F1_{hi} \cdot \frac{P_h}{\hat{P}_h}$$

where:

P_h : Projection of the total population by sex and age group in stratum h considered as a parameter for the adjustment.

\hat{P}_h : Total number of persons estimated with the theoretical factor, by sex and age group in stratum h.

$$\text{with: } \hat{P}_h = \sum_i \sum_j \frac{M_h}{M_{hi} \cdot n_h} \cdot \frac{M'_{hi}}{m_{hi}} \cdot P_{hij} = \sum_i \sum_j F1_{hi} \cdot P_{hij}$$

P_{hij} : Total number of persons (men under and over 15 years old; women under and over 15 years old) in dwelling j of the first-stage unit (section or residential block) i in stratum h.

It should be noted that the sum of all persons in stratum h, expanded by the adjusted expansion factor, results in the exact projected population for that stratum, as expected, according to the construction of this adjustment factor.

$$\sum_i \sum_j \sum_k (F2_{hi}) \cdot P_{hijk} = \sum_i \sum_j \sum_k \left(F1_{hi} \cdot \frac{P_h}{\hat{P}_h} \right) \cdot P_{hijk} = \sum_i \sum_j \sum_k \frac{F1_{hi} \cdot P_{hijk}}{\hat{P}_h} \cdot P_h = \frac{\hat{P}_h}{\hat{P}_h} \cdot P_h = P_h$$

12 Total and ratio estimators

For both estimators and their variances, we define a sampling stratum h for the basic geographic area which corresponds, in general terms, to the intersection or union of intersections of Urban and Rural strata with the communes. These were defined for the random selection and independent of the first-stage units (sections or residential blocks as clusters of houses). These result in the first disaggregated estimates. From these strata, estimate levels are constructed simply because aggregation for these levels requires a certain level of precision.

To facilitate the interpretation of estimators, variances and sums, sub-index h will refer to the geographic stratum, sub-index i will refer to the section or residential block within stratum h, the subscript j will refer to the dwelling in section i of stratum h and finally the subscript k to the person inside dwelling j of section i within stratum h.

Statistically, it is demonstrated that the final estimator used is a ratio estimator, from which totals are derived, whether quantitative or qualitative. The estimator associated to the Total corresponds to a self-weighted ratio estimator, biased but consistent²³ and that presents lower Mean Squared Error than the Total Standard Estimator²⁴.

The nomenclature used for the Estimators is:

$F1_{hi}$: Theoretical factor of expansion per person associated to section or residential block **i** within stratum **h**.

$F2_{hi}$: Factor of expansion adjusted according to population projections

$P_{hijk} = 1$, for every person **k** in dwelling **j** of section or residential block **i** in stratum **h**.

P_{hijk_c} : Person **k** in dwelling **j** of section or residential block **i** within stratum **h**, who responds to characteristic **C**.

That is:
$$P_{hijk_c} = \begin{cases} 1 & , \text{if the person respond to characteristic C} \\ 0 & , \text{in other cases} \end{cases}$$

Y_{hijk} : Value of variable **Y** measured in person **k** of dwelling **j**, in section or residential block **i**, within stratum **h**.

Y_{hijk_c} : Value of variable **Y** measured in person **k** of dwelling **j**, in section or residential block **i**, within stratum **h** who responds to characteristic **C**.

That is:
$$Y_{hijk_c} = \begin{cases} Y_{hijk} & , \text{if the person responds to characteristic C} \\ 0 & , \text{in other cases} \end{cases}$$

We can also express:

$$Y_{hijk_c} = Y_{hijk} \cdot P_{hijk_c} = \begin{cases} Y_{hijk} \cdot 1 = Y_{hijk} & , \text{If the persons responds to} \\ & \text{characteristic C} \\ Y_{hijk} \cdot 0 = 0 & , \text{in other cases} \end{cases}$$

12.1 Estimators of the Quantitative Total of variable **Y**

$\hat{Y}_h = \sum_i \sum_j \sum_k F2_{hi} \cdot Y_{hijk}$, Estimated total of variable **Y** in stratum **h**.

²³ A consistent estimator is asymptotically unbiased.

²⁴ Lohr considers the linear association between the number of employed and the total population.

$\hat{Y} = \sum_h \sum_i \sum_j \sum_k F2_{hi} \cdot Y_{hijk} = \sum_h \hat{Y}_h$, Estimated total of variable **Y** within the requested level of estimation.

The estimated total at the required level of estimation is obtained as the sum of the totals over all strata that make up that level.

$\hat{Y}_{h-c} = \sum_i \sum_j \sum_k F2_{hi} \cdot Y_{hijk-c}$, estimated total of variable **Y** in persons that respond to characteristic **C**, in stratum **h**.

$\hat{Y}_c = \sum_h \sum_i \sum_j \sum_k F2_{hi} \cdot Y_{hijk-c} = \sum_h \hat{Y}_{h-c}$, Estimated total of variable **Y** in persons that respond to characteristic **C**, within the required level of estimation.

Again, the estimated total at the required level of estimation is obtained as the sum of the Totals over all strata that make up that level.

12.2 Estimators of qualitative total or persons who respond to certain characteristic

$\hat{P}_h = \sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk} = P_h$, Estimated total of persons in stratum **h**.

The estimated total of persons in stratum **h** coincides with the population projections given as parameter.

$\hat{P} = \sum_h \sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk} = \sum_h \hat{P}_h$, Estimated total of persons in the requested level of estimation.

$\hat{P}_{h-c} = \sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk-c}$, Estimated total of persons who respond to characteristic **C** in stratum **h**.

$\hat{P}_c = \sum_h \sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk-c} = \sum_h \hat{P}_{h-c}$, Estimated total of persons who respond to characteristic **C** in the requested level of estimation.

12.3 Ratio estimators

The quotient of two variables that are estimated is often referred to as ratio. Both the numerator and denominator of this quotient can be quantitative or qualitative variables.

Again, sub index **h** will refer to the stratum, **i** to the section or residential block, **j** to the dwelling and finally **k** to the person.

12.3.1 Ratio estimator of two quantitative variables X and Y

The ratio of quantitative variable **Y** with respect to quantitative variable **X**, in stratum **h**, is calculated with the following formula:

$$\hat{R}_{h_{y/x}} = \frac{\hat{Y}_h}{\hat{X}_h} = \frac{\sum_i \sum_j \sum_k F1_{hi} \cdot Y_{hijk}}{\sum_i \sum_j \sum_k F1_{hi} \cdot X_{hijk}} = \frac{\sum_i \sum_j \sum_k F2_{hi} \cdot Y_{hijk}}{\sum_i \sum_j \sum_k F2_{hi} \cdot X_{hijk}} = \frac{\text{Estimated total of variable Y in stratum h}}{\text{Estimated total of variable X in stratum h}}$$

Note that the ratio estimator of variable Y and variable X can be calculated with the theoretical expansion factors as well as with adjusted factors of expansion.

Similarly, we obtain the ratio estimator within the required estimate level as a combined ratio.

$$\hat{R}_{y/x} = \frac{\hat{Y}}{\hat{X}} = \frac{\sum_h \sum_i \sum_j \sum_k F2_{hi} \cdot Y_{hijk}}{\sum_h \sum_i \sum_j \sum_k F2_{hi} \cdot X_{hijk}} = \frac{\text{Estimated total of variable Y in requested level}}{\text{Estimated total of variable X in requested level}}$$

12.3.2 Ratio estimator of quantitative variable Y regarding a qualitative variable

In general, in household surveys the average of quantitative total of variable **Y** (income, expenses, etc.) is calculated in respect of the total number of persons **P** or total number of persons who respond to characteristic **C**, P_c . If the ratio corresponds to quantitative

variable **Y** in respect of qualitative total **P**, we will denote this ratio as \hat{R}_{h_y} and when it is in respect of the total number of persons who respond to characteristic **C** we will denote it as

$$\hat{R}_{h_{y/c}}.$$

These ratios are therefore expressed as follows:

$$\hat{R}_{h_y} = \frac{\hat{Y}_h}{\hat{P}_h} = \frac{\sum_i \sum_j \sum_k F2_{hi} \cdot Y_{hijk}}{\sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk}} = \frac{\text{Estimated total of variable Y, in stratum h}}{\text{Estimated total of persons in stratum h}}$$

$$\hat{R}_{h_y/c} = \frac{\hat{Y}_h}{\hat{P}_{h_c}} = \frac{\sum_i \sum_j \sum_k F2_{hi} \cdot Y_{hijk}}{\sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_c}} = \frac{\text{Estimated total of variable Y, in stratum h}}{\text{Estimated total of persons who respond to characteristic C, in stratum h}}$$

$$\hat{R}_{h_yc} = \frac{\hat{Y}_{h_c}}{\hat{P}_h} = \frac{\sum_i \sum_j \sum_k F2_{hi} \cdot Y_{hijk_c}}{\sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk}} = \frac{\text{Estimated total of variable Y in persons with characteristic C, in stratum h}}{\text{Estimated total of persons, in stratum h}}$$

$$\hat{R}_{h_yc/c} = \frac{\hat{Y}_{h_c}}{\hat{P}_{h_c}} = \frac{\sum_i \sum_j \sum_k F2_{hi} \cdot Y_{hijk_c}}{\sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_c}} = \frac{\text{Estimated total of variable Y in persons of characteristic C, in stratum h}}{\text{Estimated total of persons who respond to characteristic C, in stratum h}}$$

12.3.3 Ratio estimator of two qualitative variables

The ratio estimator between persons who respond to characteristic B (employed, unemployed, inactive, etc.) and persons of characteristic C (persons over 15 years old, for example), in stratum **h**, is calculated as follows:

$$\hat{R}_{h_b/c} = \hat{T}_{h_b/c} = \frac{\hat{P}_{h_b}}{\hat{P}_{h_c}} = \frac{\sum_i \sum_j \sum_k F1_{hi} \cdot P_{hijk_b}}{\sum_i \sum_j \sum_k F1_{hi} \cdot P_{hijk_c}} = \frac{\sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_b}}{\sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_c}}$$

$$= \frac{\text{Estimated total of persons with characteristic B, in stratum h}}{\text{Estimated total of persons with characteristic C, in stratum h}}$$

As for the quantitative variables, the ratio estimator can be calculated using the theoretical factor or the adjusted factor for population projection, within each stratum.

Similarly we obtain the ratio estimator of persons with characteristic B in relation to persons with characteristic C, within the level required:

$$\hat{R}_{b/c} = \hat{T}_{b/c} = \frac{\hat{P}_b}{\hat{P}_c} = \frac{\sum_h \sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_b}}{\sum_h \sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_c}} = \frac{\text{Estimated total pf persons with characteristic B}}{\text{Estimated total of persons with characteristic C}}$$

Ratios calculated with variables related to the labour market are all persons over 15 years old, and more specifically, as these are calculated by sex, there are ratios on men o women over 15, respectively.

To make it simpler, when referring to a ratio of persons with characteristic B over total number of persons (over 15 years old) in stratum **h**, or on the required level of estimation, we will denote them as \hat{R}_{h_b} and \hat{R}_b , respectively.

$$\text{That is: } \hat{R}_{h_b} = \frac{\hat{P}_{h_b}}{\hat{P}_h} \quad \text{and} \quad \hat{R}_b = \frac{\hat{P}_b}{\hat{P}} = \frac{\sum_h \hat{P}_{h_b}}{\sum_h \hat{P}_h}$$

Note: Remember that $\hat{P}_h = P_h$, that is to say, by construction, estimate of persons within each stratum and level matches the population projection.

12.3.4 Unemployment rate estimator

The unemployment rate is a particular case of the above, where B is for unemployed (D) and C is for persons belonging to the labour force (FT), where a person that is in the labour force is either employed (O) or unemployed.

$$\hat{T}_{h_D} = \hat{T}_{h_D/FT} = \frac{\sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_D}}{\sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_FT}} = \frac{\text{Estimated total of unemployed persons in stratum h}}{\text{Estimated total of persons in the labour force, in stratum h}} = \frac{\hat{P}_{h_D}}{\hat{P}_{h_FT}}$$

We have $\hat{P}_{h_FT} = \hat{P}_{h_O} + \hat{P}_{h_D}$, where \hat{P}_{h_FT} is the estimated total of persons in the labour force; \hat{P}_{h_O} is the estimated total of employed people; and \hat{P}_{h_D} is the estimated total of unemployed persons in stratum **h**.

For the requested level of estimation, the unemployment rate is calculated as follows:

$$\hat{T}_D = \hat{T}_{D/FT} = \frac{\sum_h \sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_D}}{\sum_h \sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_FT}} = \frac{\text{Estimated total of unemployed persons in the requested level}}{\text{Estimated total of persons in the labour force, in the requested level}} = \frac{\hat{P}_D}{\hat{P}_{FT}}$$

12.3.5 Estimator of the standardised unemployment rate

The standardised unemployment rate according to OECD criteria is another ratio estimator that is calculated as follows:

$$\hat{T}_{h_D^*} = \hat{T}_{h_D^*/FT^*} = \frac{\sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_D^*}}{\sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_FT^*}} = \frac{\hat{P}_{h_D^*}}{\hat{P}_{h_FT^*}}$$

Where $\hat{P}_{h_D^*}$: Total number of unemployed plus Starters in stratum **h**

$\hat{P}_{h_FT^*}$: Total labour force plus Starters in stratum **h**

Likewise, for the level of estimation needed, the formula is as follows:

$$\hat{T}_{D^*} = \hat{T}_{D^*/FT^*} = \frac{\sum_h \sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_D^*}}{\sum_h \sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_FT^*}} = \frac{\hat{P}_{D^*}}{\hat{P}_{FT^*}}$$

Where \hat{P}_{D^*} : Total number of unemployed plus Starters in the requested level of estimation

\hat{P}_{FT^*} : Total labour force plus Starters in the requested level

12.3.6 Employment rate estimator

The employment rate is calculated using the following formula:

$$\hat{T}_{h_Oc} = \hat{T}_{h_Oc/May} = \frac{\sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_Oc}}{\sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_May}} = \frac{\text{Estimated total of employed persons in stratum h}}{\text{Estimated total of senior people in stratum h}} = \frac{\hat{P}_{h_Oc}}{\hat{P}_{h_May}}$$

Likewise, for the level of estimation needed, the formula is as follows:

$$\hat{T}_{Oc} = \hat{T}_{Oc/May} = \frac{\sum_h \sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_Oc}}{\sum_h \sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_May}} = \frac{\text{Estimated total of employed persons in the level of estimation requested}}{\text{Estimated total of seniors in the requested level of estimation}} = \frac{\hat{P}_{Oc}}{\hat{P}_{May}}$$

12.3.6 Participation rate estimator

The participation rate in the labour market is calculated using the following formula:

$$\hat{T}_{h_Part} = \hat{T}_{h_FT/May} = \frac{\sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_FT}}{\sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_May}} = \frac{\text{Estimated total of persons in the labour force, within stratum h}}{\text{Estimated total of senior persons in stratum h}} = \frac{\hat{P}_{h_FT}}{\hat{P}_{h_May}}$$

Likewise, for the level of estimation needed, the formula to calculate the participation rate is as follows:

$$\hat{T}_{Part} = \hat{T}_{FT/May} = \frac{\sum_h \sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_FT}}{\sum_h \sum_i \sum_j \sum_k F2_{hi} \cdot P_{hijk_May}} = \frac{\text{Estimated total of persons in the labour force, within the level of estimation needed}}{\text{Estimated total of senior persons in the requested level}} = \frac{\hat{P}_{FT}}{\hat{P}_{May}}$$

Equality is achieved: $Oc + D = FT$ and $FT + Inac = May$

Where:

Oc : Employed persons

D : Unemployed persons

FT : Persons in the labour force

$Inac$: Inactive persons

May : Over 15 years old

13 Variance estimation

It is worth pointing out that for purposes of calculating the variance of the estimators we use the theoretical factor, namely, the factor of expansion not adjusted according to population projections. This is to avoid the increase or decrease in the variance due to rescaling used to adjust the factors, increase or decrease that is not attributed to the variability of the labour phenomenon observed. It is also important to note that both estimator variances and ratios are calculated by sex within each stratum, using the variability between first stage units (sections and/or residential blocks) as the units contributing to variance growth.

The nomenclature used to calculate the variances is the same as that of estimators. As explained above, total estimates are made using the ratio estimator, so that expressions correspond to this estimator.

13.1 Total ratio variance of variable Y in relation to variable X

Let Y be the total of the variable to be estimated and X a variable that is linearly related to Y. Let R be the ratio of the total of variable Y in respect of the total of variable X, that is,

$R = \frac{Y}{X}$ y $\hat{R} = \frac{\hat{Y}}{\hat{X}}$ its estimator. Then to calculate the variance we use the mean square error calculation of the ratio estimator, denoted as:

$$ECM\left(\hat{R}\right) = E\left(\hat{R} - R\right)^2 = E\left(\frac{\hat{Y}}{\hat{X}} - R\right)^2 = E\left(\frac{\hat{Y} - R \cdot \hat{X}}{\hat{X}}\right)^2 \approx \frac{E\left(\hat{Y} - R \cdot \hat{X}\right)^2}{\hat{X}^2}$$

$$\text{But } E\left(\hat{Y} - R \cdot \hat{X}\right) = E\left(\hat{Y}\right) - R \cdot E\left(\hat{X}\right) = Y - R \cdot X = Y - \frac{Y}{X} \cdot X = 0$$

Therefore we can express:

$$E\left(\hat{Y} - R \cdot \hat{X}\right)^2 = E\left(\left[\hat{Y} - R \cdot \hat{X}\right] - E\left[\hat{Y} - R \cdot \hat{X}\right]\right)^2 = \text{Var}\left(\hat{Y} - R \cdot \hat{X}\right) \approx \text{Var}\left(\hat{D}\right)$$

Where $\hat{D} = \hat{Y} - R \cdot \hat{X}$ the variable whose mean is $E\left(\hat{D}\right) = 0$.

We know that the mean square error of an estimator is the sum of its variance and its squared bias (*sesgo*) – the bias being approximately equal to zero, that is to say:

$$ECM\left(\hat{R}\right) = \text{Var}\left(\hat{R}\right) + \text{Sesgo}^2\left(\hat{R}\right) \text{ with: } \text{Sesgo}\left(\hat{R}\right) = E\left(\hat{R}\right) - R \approx 0$$

Finally:

$ECM\left(\hat{R}\right) \approx Var\left(\hat{R}\right) \approx \frac{Var\left(\hat{Y} - R \cdot \hat{X}\right)}{\hat{X}^2}$, so the ratio variance estimate in

stratum **h** is represented by:

$$Var\left(\hat{R}_h\right) \doteq \frac{1}{\hat{X}_h^2} \cdot \frac{n_h}{n_h-1} \cdot \sum_i \left(\left[\hat{Y}_{hi} - \hat{R}_h \cdot \hat{X}_{hi} \right] - \frac{1}{n_h} \sum_i \left[\hat{Y}_{hi} - \hat{R}_h \cdot \hat{X}_{hi} \right] \right)^2$$

$$Var\left(\hat{R}_h\right) \doteq \frac{1}{\hat{X}_h^2} \cdot \frac{n_h}{n_h-1} \cdot \sum_i \left(\hat{D}_{hi} - \hat{D}_h \right)^2 \quad \text{with: } \hat{D}_{hi} = \hat{Y}_{hi} - \hat{R}_h \cdot \hat{X}_{hi}$$

$$\hat{Y}_{hi} = F1_{hi} \cdot Y_{hi} = F1_{hi} \cdot \sum_j Y_{hij} \quad ; \quad \hat{D}_h = \frac{1}{n_h} \cdot \sum_i \left(\hat{Y}_{hi} - \hat{R}_h \cdot \hat{X}_{hi} \right) \quad \text{and}$$

$$\hat{R}_h = \frac{\hat{Y}_h}{\hat{X}_h} = \frac{\sum_i F2_{hi} \cdot Y_{hi}}{\sum_i F2_{hi} \cdot X_{hi}}$$

Knowing that in stratum **h** $\hat{D}_h = 0$, the variance of the ratio estimator reduces to:

$$\boxed{Var\left(\hat{R}_h\right) \doteq \frac{1}{\hat{X}_h^2} \cdot \frac{n_h}{n_h-1} \cdot \sum_i \left(\hat{Y}_{hi} - \hat{R}_h \cdot \hat{X}_{hi} \right)^2}$$

13.2 Total variance of variable **Y** in stratum **h** at the required level

This is obtained directly from the estimate of the ratio variance.

Provided that $\hat{Y}_h = \hat{R}_h \cdot \hat{X}_h$, then $Var\left(\hat{Y}_h\right) = \hat{X}_h^2 \cdot Var\left(\hat{R}_h\right)$, reducing to:

$$Var\left(\hat{Y}_h\right) \doteq \frac{n_h}{n_h-1} \cdot \sum_i \left(\hat{Y}_{hi} - \hat{R}_h \cdot \hat{X}_{hi} \right)^2$$

To estimate the total of variable **Y** at the required level of estimation, we use the combined ratio of all strata that make up that level, namely:

$$\hat{R} = \frac{\hat{Y}}{\hat{X}} = \frac{\sum_h \sum_i F2_{hi} \cdot Y_{hi}}{\sum_h \sum_i F2_{hi} \cdot X_{hi}} \quad , \text{ therefore } \hat{D}_h = \frac{1}{n_h} \cdot \sum_i \left(\hat{Y}_{hi} - \hat{R} \cdot \hat{X}_{hi} \right) \neq 0$$

So the expressions of ratio and total variances remain as follows:

$$Var\left(\hat{R}\right) \doteq \frac{1}{\hat{X}^2} \cdot \sum_h \frac{n_h}{n_h-1} \cdot \sum_i \left(\left[\hat{Y}_{hi} - \hat{R} \cdot \hat{X}_{hi} \right] - \frac{1}{n_h} \cdot \sum_i \left[\hat{Y}_{hi} - \hat{R} \cdot \hat{X}_{hi} \right] \right)^2$$

$$\widehat{Var}(\widehat{Y}) \doteq \sum_h \frac{n_h}{n_h - 1} \cdot \sum_i \left(\left[\widehat{Y}_{hi} - \widehat{R} \cdot \widehat{X}_{hi} \right] - \frac{1}{n_h} \cdot \sum_i \left[\widehat{Y}_{hi} - \widehat{R} \cdot \widehat{X}_{hi} \right] \right)^2$$

An estimate of the Total variance in variable **Y**, slightly biased at the required level of estimation, is given by:

$$\widehat{Var}(\widehat{Y}) \doteq \sum_h \frac{n_h}{n_h - 1} \cdot \sum_i \left(\widehat{Y}_{hi} - \widehat{R}_h \cdot \widehat{X}_{hi} \right)^2 = \sum_h \widehat{V}(\widehat{Y}_h)$$

The previous variance is the sum of the variances of the different strata, and is the one used to calculate the variance of the Totals in household surveys, assuming that selection is independent in each stratum and also that the sample size is sufficiently large. For small samples of primary units in each stratum, it is recommended to estimate the Total with the combined ratio estimator, which reduces the bias of the variance.

For estimates of ratio at the level required, we use above combined ratio, which can also be expressed as:

$$\widehat{Var}(\widehat{R}) \doteq \frac{1}{\widehat{X}^2} \cdot \sum_h \frac{n_h}{n_h - 1} \cdot \sum_i \left(\left[\widehat{Y}_{hi} - \widehat{Y}_h \right] - \widehat{R} \cdot \left[\widehat{X}_{hi} - \widehat{X}_h \right] \right)^2$$

$$\widehat{Var}(\widehat{R}) \doteq \frac{1}{\widehat{X}^2} \cdot \sum_h n_h \cdot \left(\widehat{S}_{\widehat{Y}_h}^2 - 2 \cdot \widehat{R} \cdot \widehat{S}_{\widehat{X}_h, \widehat{Y}_h}^2 + \widehat{R}^2 \cdot \widehat{S}_{\widehat{X}_h}^2 \right), \text{ where}$$

$$\widehat{S}_{\widehat{Y}_h}^2 = \frac{1}{n_h - 1} \cdot \sum_i \left(\widehat{Y}_{hi} - \widehat{Y}_h \right)^2, \quad \widehat{S}_{\widehat{X}_h}^2 = \frac{1}{n_h - 1} \cdot \sum_i \left(\widehat{X}_{hi} - \widehat{X}_h \right)^2 \quad \text{and}$$

$$\widehat{S}_{\widehat{X}_h, \widehat{Y}_h}^2 = \frac{1}{n_h - 1} \cdot \sum_i \left(\widehat{X}_{hi} - \widehat{X}_h \right) \cdot \left(\widehat{Y}_{hi} - \widehat{Y}_h \right)$$

Almost all estimators and variances are calculated on persons belonging to the labour market, excluding minors (under 15 years old) and therefore the people considered are only over 15 years old, unless otherwise specified.

To calculate the variances for ratio and Total estimators, qualitative variables as Employed, Unemployed, belonging to the Labour Force, Unemployment Rate, Participation Rate, etc. may be used.

Additional nomenclature:

$$P_{hi} = \sum_j P_{hij} = \sum_j \sum_k P_{hijk}, \text{ Number of persons of section or residential block } \mathbf{i} \text{ in stratum } \mathbf{h}.$$

$P_{hi_c} = \sum_j P_{hij_c} = \sum_j \sum_k P_{hijk_c}$, Number of persons of section or residential block **i** in stratum **h**.

13.3 Ratio variance between persons with characteristic B and persons with characteristic C, in the sample stratum h

$$\hat{V}\left(\hat{R}_{h_b/c}\right) = \frac{n_h}{n_h - 1} \cdot \frac{\sum_i \left[F1_{hi} \cdot (P_{hi_b} - \hat{R}_{h_b/c} \cdot P_{hi_c}) - \frac{1}{n_h} \cdot \sum_i F1_{hi} \cdot (P_{hi_b} - \hat{R}_{h_b/c} \cdot P_{hi_c}) \right]^2}{\hat{P}_{h_c}^2}$$

Where $\hat{R}_{h_b/c} = \frac{\hat{P}_{h_b}}{\hat{P}_{h_c}}$ y $\hat{P}_{h_c}^2 = \left(\sum_i F2_{hi} \cdot P_{hi_c} \right)^2$ is the squared of the estimation of Total number of persons with characteristic C, in stratum **h**.

But:

$$\sum_i F1_{hi} \cdot (P_{hi_b} - \hat{R}_{h_b/c} \cdot P_{hi_c}) = \sum_i F1_{hi} \cdot P_{hi_b} - \hat{R}_{h_b/c} \cdot \sum_i F1_{hi} \cdot P_{hi_c} = \hat{P}_{h_b} - \hat{R}_{h_b/c} \cdot \hat{P}_{h_c} = 0$$

At level of stratum **h**, resulting in:

$$\hat{V}\left(\hat{T}_{h_b/c}\right) = \hat{V}\left(\hat{R}_{h_b/c}\right) = \frac{n_h}{n_h - 1} \cdot \frac{\sum_i \left[F1_{hi} \cdot (P_{hi_b} - \hat{R}_{h_b/c} \cdot P_{hi_c}) \right]^2}{\hat{P}_{h_c}^2}$$

13.4 Ratio variance between persons with characteristic B and persons with characteristic C, at the required level of estimation

$$\hat{V}\left(\hat{R}_{b/c}\right) = \frac{1}{\hat{P}_c^2} \cdot \sum_h \frac{n_h}{n_h - 1} \cdot \sum_i \left[F1_{hi} \cdot (P_{hi_b} - \hat{R}_{b/c} \cdot P_{hi_c}) - \frac{1}{n_h} \cdot \sum_i F1_{hi} \cdot (P_{hi_b} - \hat{R}_{b/c} \cdot P_{hi_c}) \right]^2$$

$$\text{Where: } \hat{P}_c = \sum_h \sum_i F2_{hi} \cdot P_{hi_c} \quad \text{y} \quad \hat{R}_{b/c} = \frac{\hat{P}_b}{\hat{P}_c} = \frac{\sum_h \sum_i F2_{hi} \cdot P_{hi_b}}{\sum_h \sum_i F2_{hi} \cdot P_{hi_c}}$$

Being the combined ratio at the required level of estimation, the average in brackets is not annulled as the average at stratum level in the expression of the variance above.

The above expression for the variance can also be written as:

$$\hat{V}\left(\hat{R}_{b/c}\right) = \frac{1}{\hat{P}_c^2} \cdot \sum_h \frac{n_h}{n_h - 1} \cdot \sum_i \left[\hat{D}_{hi} - \hat{D}_h \right]^2 \quad \text{where: } \hat{D}_{hi} = \hat{P}_{hi_b} - \hat{R}_{b/c} \cdot \hat{P}_{hi_c}$$

$$\hat{D}_h = \frac{1}{n_h} \cdot \sum_i \left(\hat{P}_{hi_b} - \hat{R}_{b/c} \cdot \hat{P}_{hi_c} \right), \quad \hat{P}_{hi_b} = F1_{hi} \cdot P_{hi_b} \quad \text{and} \quad \hat{P}_{hi_c} = F1_{hi} \cdot P_{hi_c}$$

13.5 Unemployment rate variance

A particular case of the ratio above is the unemployment rate: $\hat{T}_D = \hat{R}_{D/FT} = \frac{\hat{P}_D}{\hat{P}_{FT}} = \hat{R}_{b/c}$

Where: $b \equiv D$, Unemployed and $c \equiv FT$ belonging to the Labour Force.

\hat{P}_D is the estimated total of Unemployed at the level of estimation requested.

\hat{P}_{FT} is the estimated Total of persons in the Labour Force at the requested level.

13.6 Total variance of variable Y in stratum h at stratum h

The variable Y total is estimated, the same that qualitative totals, by means of a ratio estimator. This ratio, rather than being in respect of another quantitative variable, is in respect of a qualitative variable, such as the total number of people (or persons over 15), that is:

$$\hat{V}(\hat{Y}_h) = \frac{n_h}{n_h - 1} \cdot \sum_i \left[F1_{hi} \cdot (Y_{hi} - \hat{R}_h \cdot P_{hi}) \right]^2 \quad \text{with: } \hat{R}_h = \frac{\hat{Y}_h}{\hat{P}_h} = \frac{\sum_i F2_{hi} \cdot Y_{hi}}{\sum_i F2_{hi} \cdot P_{hi}}$$

Where:

Y_{hi} : Total of variable Y in section or residential block **i** in stratum **h**.

P_{hi} : Total number of persons in section or residential block **i** in stratum **h**.

\hat{R}_h : Ratio of the Total of variable **Y** regarding the Total number of persons in stratum **h**.

13.7 Variance of the Average per person of variable Y in stratum h

As the estimated total number of persons in stratum **h** matches population projections, meaning that $\hat{P}_h = P_h$, in the average per person, \hat{P}_h is considered a fixed parameter.

$$\hat{V}\left(\frac{\hat{Y}_h}{\hat{P}_h}\right) = \hat{V}\left(\frac{\hat{Y}_h}{P_h}\right) = \frac{\hat{V}(\hat{Y}_h)}{P_h^2}$$

13.8 Variance of Total estimated number of persons with characteristic C for stratum h

The variance of the estimate of the Total number of people with characteristic C in stratum h is obtained with the addition of all sections belonging to the stratum. The mathematical expression is as follows:

$$\hat{V}(\hat{P}_{h-c}) = \frac{n_h}{n_h - 1} \cdot \sum_i \left[F1_{hi} \cdot (P_{hi-c} - \hat{R}_{h-c} \cdot P_{hi}) \right]^2 \quad \text{with: } \hat{R}_{h-c} = \frac{\hat{P}_{h-c}}{\hat{P}_h}$$

13.9 Variance of estimated Total of persons with characteristic C for a certain level of estimation

The variance of the estimated total of persons who respond to characteristic C, for a certain level of estimation, is obtained by means of the addition of the variances of the strata that make up that level of estimation.

$$\hat{V}(\hat{P}_c) = \sum_h \frac{n_h}{n_h - 1} \cdot \sum_i \left[F1_{hi} \cdot (P_{hi-c} - \hat{R}_{h-c} \cdot P_{hi}) \right]^2 = \sum_h \hat{V}(\hat{P}_{h-c})$$

As for the quantitative variables, when the sample size of primary units in each stratum is too small, it is recommended to use the Estimator of Total obtained from a total combined ratio, that is:

$$\hat{V}ar(\hat{P}_c) \doteq \sum_h \frac{n_h}{n_h - 1} \cdot \sum_i \left(\left[\hat{P}_{hi-c} - \hat{R}_c \cdot \hat{P}_{hi} \right] - \frac{1}{n_h} \cdot \sum_i \left[\hat{P}_{hi-c} - \hat{R}_c \cdot \hat{P}_{hi} \right] \right)^2$$

$$\text{with } \hat{R}_c = \frac{\hat{P}_c}{\hat{P}} = \frac{\sum_h \sum_i F2_{hi} \cdot P_{hi-c}}{\sum_h \sum_i F2_{hi} \cdot P_{hi}} = \frac{\sum_h \sum_i F2_{hi} \cdot P_{hi-c}}{P}, \text{ being the combined ratio over}$$

all strata.

14 Coefficient of variation and error associated to the estimator

The coefficient of variation associated to an estimate indicates the relative accuracy of such estimate and is obtained from the quotient between the standard deviation (square root of the variance) and the value of the estimate. This value indicates the relative variability as if it was associated to an interval estimate of about 70%, assuming a normal behaviour of the variable under study.²⁵

²⁵ An interval estimate with 70% confidence is associated with a value "1" in the standard distribution, while an estimate with 95% confidence is linked to value "1.96".

In general, almost all interval estimates consider a 95% confidence, so we define the relative error as the double of the coefficient of variation.

14.1 Coefficient of variation and relative error

The mathematical expression of the coefficient of variation is: $Cv(\hat{Y}) = \frac{\sqrt{\hat{V}(\hat{Y})}}{\hat{Y}}$ And for the relative error we have: $E_r(\hat{Y}) = 1,96 \cdot Cv(\hat{Y})$, both expressed as percentages.

For example, that the relative error is 20% can be construed as that the real value of the variable Y in the population is about 20% of the estimated value, with 95% confidence. In terms of the coefficient of variation, this would be interpreting that the real value of variable Y in the population is approximately 10%, but with 70% confidence only.

Once this value is known, it is possible to determine with the estimate, the confidence interval of the real value of variable Y. To do this, we suggest an interval with 95% confidence ($Z_k = 1.96$).

Given the sample size, whether the estimate is acceptable or not depends on the relative error or coefficient of variation obtained and of the level of estimation. In general, this error can be classified as reference, as described in the following table.

Table 3 - Relative error classification (95%) and coefficient of variation

Classification	Relative error range		Coefficient of variation range	
Very good	10%			5%
Good	10%	30%	5%	15%
Acceptable	30%	50%	15%	25%
Not recommended	50%		25%	

For example, in average, regional levels of estimation have relative errors of some 10%, just like big cities, while at national level it drops to 3.3% (see Appendix).

14.2 Absolute error and confidence interval

For the 95% confidence interval for variable Y estimate we use the absolute error (error of estimation that is expressed in the same units as the variable Y under study).²⁶

The absolute error associated with a 95% confidence estimate can be construed as the maximum difference expected between the estimated value and the real value of variable Y, with a 95% confidence estimate.

The absolute error is defined as: $E_A(\hat{Y}) = 1,96 \cdot \sqrt{\hat{V}(\hat{Y})}$

²⁶ When variable Y under study is a percentage or ratio, the absolute error is expressed in percentage points

In probability terms, it can be expressed as follows:

$Probabilidad[|\hat{Y} - Y| \leq E_A(Y)] = 95\%$ where \hat{Y} and Y are the estimate and real values in the population, respectively.

This is equivalent to saying that the 95% confidence interval for the real value of Variable Y in the population, is given by:

$$IC(Y)_{(95\%)} = [\hat{Y} - E_A(\hat{Y}); \hat{Y} + E_A(\hat{Y})]_{(95\%)} = [\hat{Y} - 1,96 \cdot \sqrt{\hat{Y}}; \hat{Y} + 1,96 \cdot \sqrt{\hat{Y}}]_{(95\%)}$$

The previous interval can also be represented in terms of the relative error:

$$IC(Y)_{(95\%)} = [\hat{Y} - E_R(\hat{Y}) \cdot \hat{Y}; \hat{Y} + E_R(\hat{Y}) \cdot \hat{Y}]_{(95\%)} = [\hat{Y} \cdot (1 - E_R(\hat{Y})); \hat{Y} \cdot (1 + E_R(\hat{Y}))]_{(95\%)}$$

where $E_A(\hat{Y}) = E_R(\hat{Y}) \cdot \hat{Y}$

APPENDIX: Sampling size and errors by level of estimationⁱ.

REGION	LEVEL OF ESTIMATION	AVERAGE UNEMPLOYMENT RATE (FIVE QUARTERS)	THEORETICAL SAMPLE SIZE	RELATIVE ERROR	ABSOLUTE ERROR
COUNTRY	NATIONAL	9.1%	35,719	3.3%	0.30%
COUNTRY	URBAN	9.9%	28,717	3.6%	0.4%
COUNTRY	GRANDES CENTROS	9.9%	23,573	4.1%	0.4%
COUNTRY	RURAL	5.6%	7,002	8.5%	0.5%
1	TARAPACÁ REGION	7.6%	1,026	18.1%	1.4%
1	IQUIQUE CITY	6.9%	564	26.4%	1.8%
1	ALTO HOSPICIO CITY	9.3%	222	31.1%	2.9%
2	ANTOFAGASTA REGION	6.8%	1,050	20.5%	1.4%
2	CALAMA CITY	6.9%	384	29.1%	2.0%
2	ANTOFAGASTA CITY	6.6%	504	33.5%	2.2%
3	ATACAMA REGION	8.6%	999	15.9%	1.4%
3	COPIAPO CITY	8.6%	438	25.1%	2.2%
3	VALLENAR CITY	9.5%	180	30.5%	2.9%
4	COQUIMBO REGION	6.8%	2,193	13.3%	0.9%
4	URBAN	8.0%	1,233	15.8%	1.3%
4	COQUIMBO CITY	8.2%	279	31.1%	2.5%
4	LA SERENA CITY	8.5%	306	31.8%	2.7%
4	OVALLE CITY	4.8%	297	36.3%	1.7%
4	ELQUI PROVINCE	5.8%	1,071	19.9%	1.2%
4	LIMARI PROVINCE	3.7%	726	27.5%	1.0%
4	CHOAPA PROVINCE	15.2%	396	17.2%	2.6%
4	RURAL	5.3%	960	23.1%	1.2%
5	VALPARAÍSO REGION	11.3%	5,055	7.5%	0.8%
5	URBAN	11.6%	4,575	7.9%	0.9%
5	RURAL	7.7%	480	24.7%	1.9%
5	PETORCA PROVINCE	12.4%	216	24.0%	3.0%
5	LOS ANDES PROVINCE	5.6%	348	33.8%	1.9%
5	LOS ANDES CITY	13.7%	462	21.9%	3.0%
5	SAN FELIPE PROVINCE	8.7%	396	24.9%	2.2%
5	QUILLOTA PROVINCE	16.0%	531	21.6%	3.4%
5	VALPARAISO PROVINCE	9.5%	1,209	18.0%	1.7%
5	MARGA MARGA PROVINCE	11.3%	879	17.6%	2.0%
5	VALPARAISO CITY	9.7%	639	24.2%	2.3%
REGION	LEVEL OF ESTIMATION	AVERAGE UNEMPLOYMENT	THEORETICAL SAMPLE SIZE	RELATIVE ERROR	ABSOLUTE ERROR

		RATE (FIVE QUARTERS)			
5	VIÑA DEL MAR CITY	8.4%	648	26.2%	2.2%
5	QUILPUE CITY	10.0%	384	30.1%	3.0%
5	VILLA ALEMANA CITY	12.6%	399	23.8%	3.0%
5	SAN ANTONIO PROVINCE	13.1%	600	20.3%	2.7%
5	GREATER VALPARAISO AREA	10.3%	1.941	13.5%	1.4%
5	SAN ANTONIO CITY	14.0%	537	21.4%	3.0%
5	QUILLOTA	16.4%	702	18.3%	3.0%
6	VI REGION DEL LIBERTADOR GENERAL BERNARDO O'HIGGINS	7.2%	2.259	12.6%	0.9%
6	URBAN	7.7%	1.389	16.1%	1.2%
6	RURAL	6.3%	870	20.0%	1.2%
6	RANCAGUA CITY	8.1%	609	27.1%	2.2%
6	SAN FERNANDO CITY	7.3%	288	40.0%	2.9%
6	CACHAPOAL PROVINCE	7.6%	1.095	18.0%	1.4%
6	COLCHAGUA PROVINCE	6.4%	681	23.0%	1.5%
6	CARDENAL CARO PROVINCE	7.1%	483	26.8%	1.9%
7	MAULE REGION	9.6%	2.456	10.5%	1.0%
7	URBAN	11.3%	1.616	11.9%	1.4%
7	RURAL	6.3%	840	20.9%	1.3%
7	CURICO CITY	8.4%	240	35.7%	3.0%
7	TALCA CITY	14.2%	408	21.1%	3.0%
7	LINARES CITY	12.8%	375	23.5%	3.0%
7	TALCA PROVINCE	14.0%	663	17.5%	2.5%
7	LINARES PROVINCE	10.6%	765	17.3%	1.8%
7	CAUQUENES PROVINCE	6.9%	344	30.4%	2.1%
7	CURICO PROVINCE	5.7%	684	23.7%	1.3%
8	BÍO BÍO REGION	12.7%	4.920	7.5%	1.0%
8	URBAN	13.5%	4.365	7.8%	1.0%
8	RURAL	6.6%	555	27.7%	1.8%
8	CHILLAN CITY	14.5%	441	20.6%	3.0%
8	LOTA CITY	17.6%	522	17.0%	3.0%
8	CORONEL CITY	19.8%	567	22.7%	4.5%
8	LOS ANGELES CITY	8.5%	369	30.5%	2.6%
8	ÑUBLE PROVINCE	13.9%	618	18.2%	2.5%
8	CONCEPCION PROVINCE	13.2%	2.458	11.3%	1.5%
8	CONCEPCION CITY	8.4%	1.045	20.9%	1.8%
8	TALCAHUANO CITY	11.3%	393	23.6%	2.7%
8	ARAUCO PROVINCE	12.3%	495	22.0%	2.7%
8	BIO-BIO PROVINCE	11.1%	996	16.7%	1.9%
8	GREATER CONCEPCION AREA	8.2%	1.120	20.1%	1.7%
REGION	LEVEL OF ESTIMATION	AVERAGE UNEMPLOYMENT RATE (FIVE	THEORETICAL SAMPLE SIZE	RELATIVE ERROR	ABSOLUTE ERROR

		QUARTERS)			
9	LA ARAUCANÍA REGION	9.7%	2.040	12.4%	1.2%
9	URBAN	11.2%	1.365	13.5%	1.5%
9	RURAL	6.8%	675	26.6%	1.8%
9	ANGOL CITY	10.6%	423	26.1%	2.8%
9	TEMUCO CITY	10.2%	606	19.8%	2.0%
9	MALLECO PROVINCE	10.2%	954	17.9%	1.8%
9	CAUTIN PROVINCE	9.3%	981	18.6%	1.7%
10	LOS LAGOS REGION	4.2%	2.349	18.8%	0.8%
10	URBAN	4.9%	1.239	22.6%	1.1%
10	RURAL	3.5%	1.110	31.2%	1.1%
10	OSORNO CITY	6.5%	519	34.8%	2.3%
10	PUERTO MONTT CITY	3.8%	444	39.3%	1.5%
10	OSORNO PROVINCE	5.3%	1.020	27.3%	1.4%
10	LLANQUIHUE PROVINCE	4.0%	846	28.1%	1.1%
10	CHILOE-PALENA PROVINCE	2.4%	483	52.5%	1.3%
11	AYSÉN DEL GENERAL CARLOS IBAÑEZ DEL CAMPO REGION	5.5%	828	24.7%	1.3%
11	COYHAIQUE CITY	5.8%	411	36.4%	2.1%
11	PUERTO AISEN CITY	8.8%	183	34.0%	3.0%
12	MAGALLANES, ANTÁRTICA CHILENA REGION	3.7%	321	44.7%	1.7%
12	PUNTA ARENAS CITY	3.3%	207	58.0%	1.9%
13	METROPOLITAN REGION	9.2%	7.991	7.0%	0.6%
13	URBAN	9.3%	7.763	7.1%	0.7%
13	RURAL	5.6%	375	28.0%	1.6%
13	SANTIAGO PROVINCE	8.4%	5.054	9.76%	0.82%
13	GREATER SANTIAGO AREA	9.0%	6.386	8.3%	0.7%
13	PUENTE ALTO CITY	9.4%	582	23.4%	2.2%
13	SAN BERNARDO CITY	13.8%	603	21.8%	3.0%
13	MELIPILLA CITY	12.6%	333	23.9%	3.0%
13	COLINA CITY	10.6%	321	27.0%	2.9%
13	CHACABUCO	8.5%	213	35.0%	3.0%
13	CORDILLERA PROVINCE	9.4%	657	22.1%	2.1%
13	MAIPO PROVINCE	12.4%	744	19.2%	2.4%
13	MELIPILLA PROVINCE	11.1%	465	20.3%	2.2%
13	TALAGANTE PROVINCE	9.2%	537	20.2%	1.9%
14	LOS RÍOS REGION	8.2%	1.092	17.7%	1.5%
14	URBAN	9.7%	705	20.4%	2.0%
14	RURAL	5.5%	387	33.8%	1.9%
14	VALDIVIA CITY	9.8%	501	24.5%	2.4%
14	VALDIVIA PROVINCE	8.2%	1.092	17.7%	1.5%
15	ARICA AND PARINACOTA REGION	8.1%	1.140	18.0%	1.5%
15	ARICA CITY	9.4%	825	19.8%	1.9%

Bibliography

- C. J. Skinner, D. Holt y T. M. Smith (1989), “Analysis of Complex Surveys.”
- Carl-Erik Sarndal, Bengt Swensson, Jan Wretman (1992), “Model Assisted Survey Sampling.”
- Cochran W. G. (1998), “Sampling techniques.”
- Des Raj (1968), “Sampling Theory.”
- Sharon L. Lohr, “Sampling: Design and Analysis.”

ⁱ The table on sampling size and errors by level of estimation has been updated on November 2, 2010.