universidade de aveiro

universidade de aveiro

Proceedings of the
19th European Conference on Research Methodology for Business and Management Studies

A Virtual Conference hosted by
University of Aveiro
Portugal
18-19 June 2020

Edited by
Dr. Manuel Au-Yong-Oliveira
Professor Carlos Costa

acpi

A conference managed by ACPI, UK

Proceedings of the

19th European Conference on Research Methods

ECRM 2019

18-19 June 2020

Hosted By

The University of Aveiro, Portugal

Edited by

Dr. Manuel Au-Yong-Oliveira and Professor Carlos Costa

Review Process
Papers submitted to this conference have been double-blind peer reviewed before final acceptance to the conference. Initially, abstracts were reviewed for relevance and accessibility and successful authors were invited to submit full papers. Many thanks to the reviewers who helped ensure the quality of all the submissions.

Ethics and Publication Malpractice Policy
ACPIL adheres to a strict ethics and publication malpractice policy for all publications – details of which can be found here: http://www.academic-conferences.org/policies/ethics-policy-for-publishing-in-the-conference-proceedings-of-academic-conferences-and-publishing-international-limited/

Self-Archiving and Paper Repositories
We actively encourage authors of papers in ACPIL conference proceedings and journals to upload their published papers to university repositories and research bodies such as ResearchGate and Academic.edu. Full reference to the original publication should be provided.

Conference Proceedings
The Conference Proceedings is a book published with an ISBN and ISSN. The proceedings have been submitted to a number of accreditation, citation and indexing bodies including Thomson ISI Web of Science and Elsevier Scopus.

Author affiliation details in these proceedings have been reproduced as supplied by the authors themselves.

The Electronic version of the Conference Proceedings is available to download from DROPBOX https://tinyurl.com/ecrm20 Select Download and then Direct Download to access the Pdf file. Free download is available for conference participants for a period of 2 weeks after the conference.

The Conference Proceedings for this year and previous years can be purchased from http://academic-bookshop.com

E-Book ISBN: 978-1-912764-60-0
E-Book ISSN: 2049-0976
Book version ISBN: 978-1-912764-59-4
Book Version ISSN: 2049-0968

Published by Academic Conferences and Publishing International Limited
Reading, UK
+44-118-972-4148
www.academic-conferences.org
info@academic-conferences.org

Proposed Representative Sampling Methodology

Milton F. Barragán-Landy[1,2], Sérgio D. Sousa[2], Fernando Romero[2] and Celina P. Leão[2]
[1] Department of Applied Chemistry and Production Systems, Faculty of Chemical Sciences, University of Cuenca, Cuenca, Ecuador.
[2] ALGORITMI Research Centre, University of Minho, Guimarães, Portugal
mfrancisco.barraganl@ucuenca.edu.ec
sds@dps.uminho.pt
fromero@dps.uminho.pt
cpl@dps.uminho.pt

Abstract: A representative sample is a subset of a population which ensures that those characteristics of the population which are under analysis are represented as completely as possible. There are different ways of estimating a representative sample of a population. When different and important characteristics of context are relevant (i.e. region, city, sectors by economic activities, economically active population and others), the selection of an appropriate method to collect data must be carefully planned. The research questions that motivated this work were: What method could help define a sector of study in a country? What criteria must be considered to define a sampling method? What is the adequate sample size considering subsectors within the sector of study? Is there any difference between sampling techniques used to define representative samples in a sector? The aim of this document is to establish an interactive and sequential process to select a sampling method and to apply it to define a sample of companies in the construction sector in Ecuador. The methodology was based on four phases; a characterization of the region and the sectors in the country; a population characterization, based on selected parameters; a sampling process based on a literature review; and a comparative analysis between two sampling techniques. A simple random sampling and a stratified random sampling within a selected sector were compared. Based on the obtained samples, 18.46% on average of the data were equally selected, independently of the sampling method. This reveals that the two methods, which answer specific objectives, can be used together to select a smaller sample. This methodology can be applied in business and management studies in other sectors and can provide an economy in terms of the resources needed for data collection without compromising the sample representativeness.

Keywords: representative sample, sampling method, context, research methodology.

1. Introduction

Sampling methodology has been a concern to the research community. According to Israel (2008) it is important to define a representative sample so that it is possible to infer to a population any behaviour that is determined in the sample. A representative sample is a subset of a population which ensures that those characteristics of the population which are under analysis are represented as completely as possible. As such, it is necessary to establish an adequate approach in the selection of the sample.

The Burnam and Koegel (1988) work is an example of a study that is concerned with the representativeness of samples of a population that meet specific criteria. The authors identified several necessary stages of research that must be considered to design and select a representative sample that is able to characterize the population. Despite the difficulties in implementing sampling designs, the authors highlighted the importance of conducting careful research and argue that studies without a representative sample must be prudently interpreted.

Sampling is a process that defines the number of cases that are selected from a certain population under study. One of the aims of the sampling process is to reduce the number of cases that can be confidently used to make inferences about a population or to build generalizations concerning an existing theory (Taherdoost, 2016). Sampling is an essential component of most research studies, allowing researchers to answer the research question by defining the data that must be collected. The type of data and information to be collected, and how the data is to be analysed, are concerns that prevent a good research design having distortion factors which affect the result of the study. To do this, and since for most situations analysing the entire population is prohibitively costly and time-consuming, a sample should be representative of the population. It is also important to select the population according to the corresponding research question. As Ramsey and Hewitt (2005) have pointed out, “a sample that is representative for a specific question is most likely not representative for a different question”.

Taherdoost (2016) describes the possible stages that can be followed when conducting any sampling methodology: (1) a clear definition of the target population under study; (2) select a sampling frame representative of the population; (3) choose a sampling technique (probability or random sampling and non-

probability or non-random sampling) according to the nature of the research design; (4) calculate the sample size based on existing formulas; (5) collect data, and finally (6) assess response rates, since each non response is likely to bias the final sample.
The present paper, through the analysis of a specific sector in a city in Ecuador, defines a four phases sampling process and six specifics results. This four phases framework simplifies the Taherdoost (2016) framework by integrating some of the six stages, as the following section will describe in detail, and thus enables a comparative exercise. The purpose of this document is to establish an interactive and sequential sampling process to select a sampling method and to determine the sample of organizations in the construction sector in Cuenca city Ecuador. This process can also be used to select a sample in other sectors and countries. A comparative analysis between two sampling techniques (simple random sampling and stratified random sampling) will be presented, emphasizing the importance and adequacy for the purpose of the research. In section 2 the collected data will be characterized and the phases of the sampling process will be defined. In section 3 the findings of the research will be presented and discussed. Finally, section 4 summarizes the results and address some conclusions and future research recommendations.

2. Methodology

In this section, the characterization of the city and sector is initially made in order to clarify the meaning of sampling. Then, the various phases that constitute the sampling process are presented.

2.1 Characterization of the city and sector from which the data was collected

Cuenca city is located in the province of Azuay in southern Ecuador (Figure 1). It has 603,269 inhabitants, with a population growth of 15% in the last seven years and it is expected that by 2020 it will have 636,996 inhabitants (INEC, 2017).

Figure 1: Cuenca location

In terms of Gross Value Added (GVA), Cuenca city represents 87.90% of all sectors of the province of Azuay (BCEcuador, 2015), which shows its economic importance. According to BCEcuador (2015), Cuenca city has a significant contribution in terms of total GVA in the Construction Sector (CS), representing 96.60% of the province of Azuay, and in Ecuador it has had a rapid growth according to data from the Central Bank of Ecuador (BCEcuador, 2015). The trajectory of the sector reflects an increase in competitiveness and a preoccupation with customer satisfaction and loyalty. According to the International Standard Industrial Classification of All Economic Activities (United Nations, 2008) and to the "Servicio de Rentas Internas" (SRI, 2017), there is a total of 2049 organizations operating in the CS.

2.2 The Sampling Process

This section describes the various phases that constitute the sampling method (Figure 2) and that resulted in the selection of the above mentioned city and sector.

Phase 1. This phase presents an overview of the situation in Ecuador in relation to economic indicators, particularly GVA, through the analysis of the country, region, planning zone (includes several provinces), province, and finally the city. This phase is important to define and recognise the population under study. Different indicators and levels would be dependent on the research question and on the availability of statistics. Compared to Taherdoost et al. (2016), phase 1 encompasses the first two stages: definition of the target population and sampling frame definition. The latter is the most difficult one (Ramsey and Hewitt, 2005).

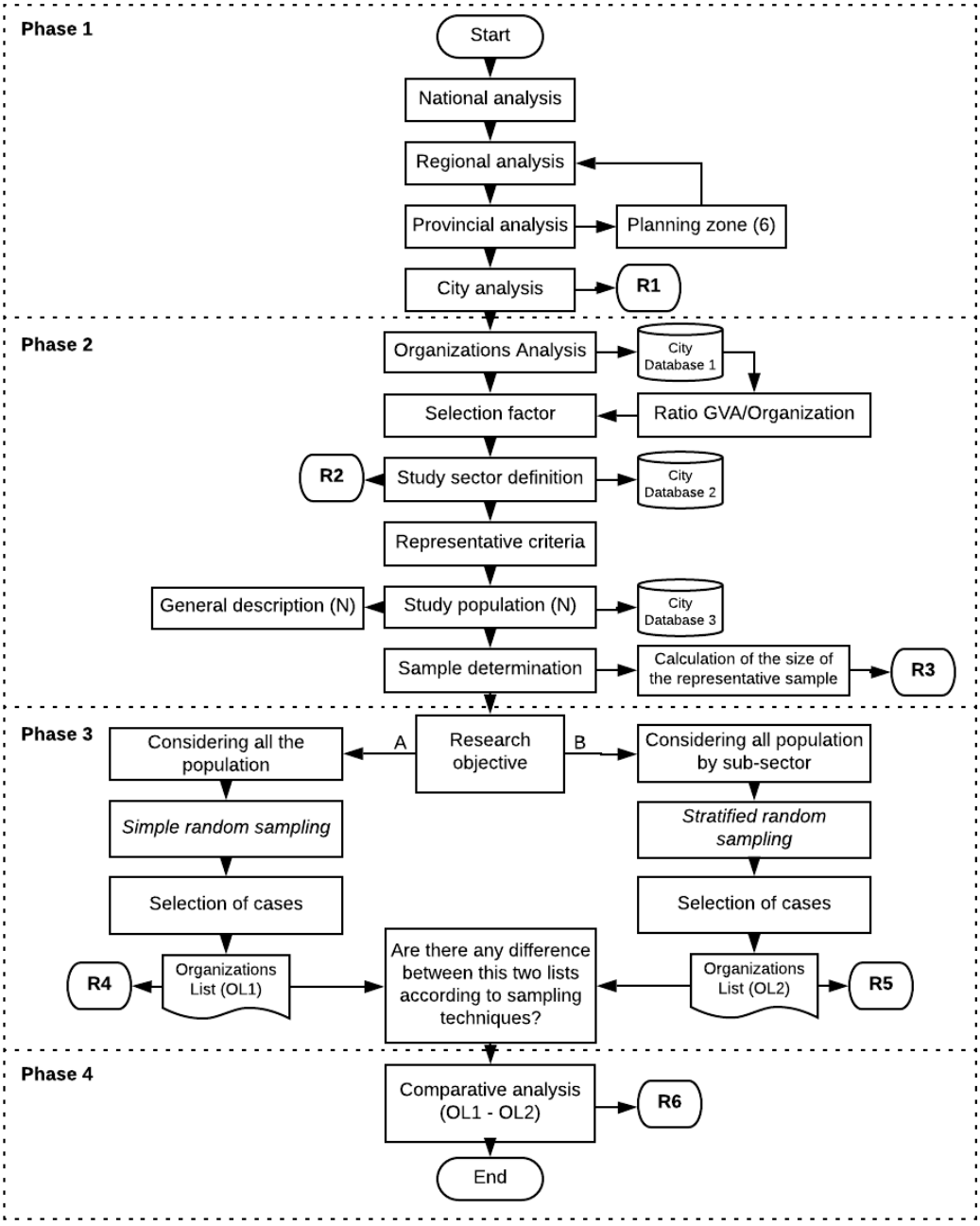

Figure 2: Flowchart to define the sampling methodology/procedure

Phase 2. The second phase corresponds to the analysis of the organizations that are operating at the city level in Cuenca city. This phase involves several steps, which start with the preparation and definition of a database of all the organizations of the city (database 1); then, the ratio between the GVA and the organizations is used to define the study sector (database 2). It is also necessary the selection and implementation of a representative criteria to define a study population (database 3).

When the study population is defined, a general description of the population is made. Then it is necessary to calculate the sample size, which can be done using different approaches/formulas. Equation (1) can be used to calculate the representative sample size for proportion (Israel, 2013; Saunders, Lewis, & Thornhill, 2016):

$$n_o = \frac{z^2 p(1-p)}{e^2} \quad (1)$$

where: n_o is the sample size, $z_{1-\alpha}$ is the value corresponding to the desired confidence level (z=1.96 for α=0.05), p is the estimated proportion of an attribute that is present in the population (usually equal to 50% as an

estimative of *p* giving the maximum sample size), and *e* is the desired level of precision (the risk the researcher is predisposed to accept).

When the population size is known (finite population), the value obtained for the sample size can be corrected:

$$n = \frac{n_o}{1+\frac{(n_o-1)}{N}} \quad (2)$$

where: n is the corrected value of the sample size, n_o is the sample size and *N* is the population size.

Phase 3. The purpose of the present work is to compare different sampling techniques, which are: (1) a simple random sampling technique is applied to all the population to select the cases and obtain the organizations list OL1; and (2) a stratified random sampling technique is applied, considering the population divided by sub-sectors, to select the cases in each sub-sector and obtain the organizations list OL2.

Phase 4. Shows a comparative analysis of the results between the two lists OL1 and OL2.

2.3 Population Characterization

When characterizing the population, there is a need to identify its dimensions related to the research objectives. Ecuador is a country located in South America, occupying a total area of 283,561 km^2 with a population of 16,689,847 inhabitants (INEC, 2018). It is divided into four regions: coast region, sierra region, amazon region and the insular region (Galápagos Islands).

The ISIC of United Nations (2008), rev.4, classifies economic activities into 21 categories (coded from A to U). For the purposes of this work, the most important indicators were Gross Domestic Product (GDP) and GVA. Some databases in the Central Bank of Ecuador are of open access, and some of them were used to collect and present data regarding the main objective of this project.

Table 1 shows the national accounts information series of the GVA, where GVA contribution to the GDP of 13.16% corresponds to the ISIC code C (Manufacturing Industries), 11.54% corresponds to L, M, N, (Real estate, Business and Rental Activities), and 10.31% to G (Wholesale and retail trade), thus identifying the sectors that contribute most to GVA.

Table 1. National Accounts Information Series – GVA (thousands of US dollars). Source: BCEcuador, 2016.

ISIC rev.4	Economic Activity/Sector	Average (2008-2016)	GVA contribution to GDP by sector
C	Manufacturing Industries	8,298,238	13.16%
L, M, N	Real Estate, Business and Rental Activities	7,271,754	11.54%
G	Wholesale and Retail Trade	6,500,494	10.31%
B	Mining and Quarrying	6,362,707	10.09%
H, J	Transportation, Information and Communications	6,306,984	10.01%
P, Q, R, S	Services to Homes	6,268,817	9.94%
F	Construction	5,768,418	9.15%
A	Agriculture, Forestry and Fishing	5,591,920	8.87%
O	Public Administration and Defence; Compulsory Social Security	3,850,344	6.11%
K	Financial and Insurance Activities	1,875,245	2.97%
D, E	Supply of Electricity and Water	1,311,523	2.08%
I	Accommodation and Food Service Activities	1,102,273	1.75%
T	Private Households with Domestic Service	178,469	0.28%
	Gross Value Added of industries (GVA)	**60,687,186**	
	Other Elements of GDP	2,350,492	3.73%
	Gross Domestic Product (GDP)	**63,037,678**	

Table 2 shows the same situation but relatively to Ecuador regions. There is an important contribution to the GVA of the Sierra Region (48.03%) followed by Coast Region (44.14%), Amazon Region (7.61%) and finally, the Insular Region (0.22%). Also, there is an important contribution to the ISIC code C-J (Manufacturing Industries and Information and Communication), followed by F (Construction and G Wholesale and Retail Trade).

Table 2. National Accounts Information Series–Regions GVA (thousands of US dollars). Source: BCEcuador, 2016.

ISIC rev.4	Amazon Region	Coast Region	Insular Region	Sierra Region	Total (GVA)
A	319,900	5,908,793	14,414	3,144,650	9,387,757
B	4,285,512	276,514	-	128,682	4,690,708
C-J	194,066	7,347,198	2,771	6,777,122	14,321,157
D-E	51,676	639,645	2,130	815,223	1,508,674
F	360,254	5,250,732	19,536	5,494,897	11,125,419
G	252,848	5,489,626	22,373	4,453,380	10,218,227
I	65,341	986,225	22,379	1,009,112	2,083,057
H-J	280,166	2,866,103	31,535	3,579,213	6,757,017
K	44,441	929,614	2,673	2,187,992	3,164,720
L-M-N	231,497	4,437,307	44,463	6,928,714	11,641,981
O	388,118	1,564,266	25,013	4,682,133	6,659,530
P	331,294	2,604,260	8,106	2,294,578	5,238,238
Q	159,763	1,401,246	3,808	1,685,863	3,250,680
R-S-T-U	35,011	928,806	7,793	1,023,732	1,995,342
Total	**6,999,887**	**40,630,335**	**206,994**	**44,205,291**	**92,042,507**
%	**7.61 %**	**44.14 %**	**0.22 %**	**48.03 %**	

The planning zone 6 is formed by provinces of Morona Santiago, Azuay and Cañar (SENPLADES, 2018). Table 3 shows the important contribution of Azuay (76.33%), followed by Cañar (16.42%) and Morona Santiago (7.25%) in the province's economy. In this zone, there is an important contribution to GVA of the ISIC code F followed by the code C-J.

Cuenca city represents the largest contribution of GVA in the Azuay Province, with 87.90%, followed by Sevilla de Oro (4.06%), Gualaceo (1.84%), Paute (1.51%) and the remaining cities. Thus, Cuenca city has an interesting economic dynamic in the province. In Azuay province, C-J were the ISIC code sectors with the largest contribution to GVA, followed by the ISIC code F.

Table 3. National Accounts Information Series – Province Accounts – GVA (thousands of US dollars). Source: BCEcuador, 2016.

ISIC rev.4	Morona Santiago	Azuay	Cañar	Total Zone 6
A	36,777	149,358	142,338	328,473
B	51	47,060	1,844	48,955
C-J	9,910	840,606	85,732	936,248
D-E	10,878	258,834	10,551	280,263
F	55,848	779,317	196,054	1,031,219
G	34,950	493,985	114,315	643,250
I	12,149	77,855	21,060	111,064
H-J	37,608	393,351	148,186	579,145
K	12,278	317,294	57,238	386,810
L-M-N	47,622	648,317	40,657	736,596
O	75,670	226,001	69,607	371,278
P	64,970	252,451	76,056	393,477
Q	48,344	227,308	55,446	331,098
R-S-T-U	5,830	53,970	6,209	66,009
Total	**452,885**	**4,765,707**	**1,025,293**	**6,243,885**
%	**7.25%**	**76.33%**	**16.42%**	

Cuenca city has important contributions from economic activities: 20% corresponds of GVA to the code C-J (Manufacturing Industries and Information and Communication), followed by 18% that corresponds to the code F (Construction), and 15 % of code L, M, and N (Real estate, Business and Rental Activities). Additionally, the

contribution of GVA to GDP of Construction in terms of the whole country is: 5.97% from Cuenca city, 6.80% from Azuay province, 8.90% from Zone 6, 63.04% from Sierra Region relative to Ecuador GDP. By considering the GVA as a pivot to compare the contribution, Cuenca city contribution to Construction GVA is 96.60% within Azuay province, 73.01% within the planning zone 6, 13.70% within Sierra Region and finally 11.01% within GDP of Ecuador. These values justify the choice of Cuenca city as the city to be sampled (R1 Figure 2).

3. Presentation and discussion of results

After identifying the population and the sampling frame (Phase 1), the next step is to analyse the organizations that operate in Cuenca city (phase 2). In a way, the reasons for choosing the city as the sampling frame (R1, end of phase 1), has led us to identify the sector to be studied (R2, phase 2, Figure 2): ISIC code F (Construction), a sector that includes general and specialized construction activities for buildings and civil engineering works (United Nations, 2008, 2015).

3.1 Representative Sample Definition

Once defined the city and the sector to be studied, the next step is to define the representative sample of dimension *n* (R3, end of Phase 2, Figure 2).

Cuenca city has a total of 92,384 organizations that operate normally and are active (SRI, 2017). An important aspect is that 26,286 organizations (28.45%) corresponds to the ISIC code G (Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles) showing that the economy of Cuenca city has a large commercial component. There are also 11,899 organizations (12.88%) in M (Professional, Scientific and Technical activities) and 10,057 organizations (10.89%) in C (Manufacturing). These three ISIC codes represent 52.22% of the organizations in Cuenca city.

Although Cuenca city has many companies in the commercial sector, it cannot be said with certainty that the organizations belonging to the sector generate higher GVA. Therefore, in Table 4 the ratio GVA/organizations is presented showing the sectors with the largest ratio.

Table 4. National Accounts Information Series – Cuenca city – GVA / selection factor (thousands of US dollars). Source: BCEcuador, 2016.

ISIC rev.4	GVA – Cuenca	Number of organizations	Ratio: GVA/Organizations
D-E	71,044	78	910.82
O	185,407	332	558.45
K	281,755	565	498.68
F	752,851	2049	367.42
B	42,217	182	231.96
P	183,636	1955	93.93
C	828,472	10057	82.38
Q	200,229	3699	54.13
A	79,421	1976	40.19
H-J	360,868	9723	37.11
L-M-N	608,539	21387	28.45
G	475,201	26286	18.08
I	73,772	4630	15.93
R-S-T-U	45,508	8553	5.32
V-W-X	-	912	-
Total	**4,188,920**	**92,384**	**45.34**

To select the most important sector, the criteria could be the one with the highest ratio GVA/organization, which is ISIC code D-E (Electricity, Gas, Steam and Air Conditioning supply, Water supply; Sewerage, Waste management and remediation activities). However, because many government companies belong to this sector, the access to detailed information has many restrictions. There is a similar situation for ISIC O (Public administration and defence; Compulsory social security) and K (Financial and Insurance activities) where access to information faces many restrictions. On the other hand, ISIC F (CS), in the 4th position (Table 4), plays a major

role in the development of the city, zone, and region, and the access to information is probably much easier (R2, Figure2).
Karuppusami and Gandhinathan (2006) state that "the quality tool Pareto analysis was used to sort and arrange the critical success factors according to the order of criticality", and this could help to redefine the 2049 organizations that currently exist in the CS in Cuenca city, as it is shown next.
Figure 3 presents the analysis of the organizations in the CS disaggregated by 37 sub-codes. The sub-code F410010 (Construction of all types of residential buildings: individual family houses, multi-family buildings, including buildings of elevated heights, housing for elderly persons, houses for beneficence, orphans, jails, quarters, convents, religious houses. Includes remodelling, renewal or rehabilitation of existing structures) represents 48.85% (1001 organizations) and the following sub-code F432102 (Installation of lighting systems, fire alarm systems, burglar alarm systems) represents 9.08% (186 organizations). These two sub-codes, together with the following three sub-codes F433020 (Installation of doors except automatic and swivel, windows, frames of doors and windows, installation of kitchen accessories, built-in wardrobes, ladders, furniture stores and similar wood or other materials, interior finishes like ceilings, wooden cover of walls, mobile bulkhead, etc.) that represent 7.56 (155 organizations), F432101 (Installation of electrical accessories, telecommunication lines, computer networks and cable television lines, including optical fibre lines, parabolic antennas, includes also connection of electrical appliances, domestic equipment and radiant heating systems) that represent 7.37% (151 organizations) and sub-code F439020 (Rental of cranes with operator and other construction equipment that cannot be assigned to a specific construction type with operator) 6.88% (141 organizations), account for the major part of the ISIC code Construction (≈80%). Thus, the total population of study can be defined by these 1634 organizations, where the most important areas of interest are represented.

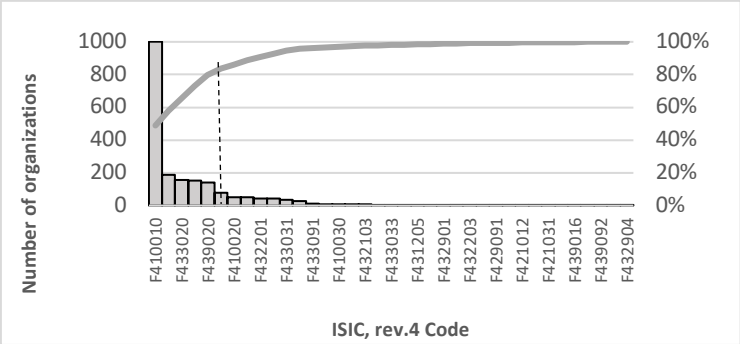

Figure 3: Pareto Diagram – Organizations in Cuenca city by ISIC, rev4, Code F (Construction).

The population of 1634 organizations has the following characteristics: average years of activity is 8.12 years, with a minimum of 0.83 years and maximum of 61.15 years; an average number of 1.10 establishments by each organization with a maximum of 29 establishments. Further influential factors are the taxpayer class and type (societies, individuals), type of accounting (internal or external) and type of regions (rural or urban).
At this point, it is possible to estimate the representative sample size, *n*, based on equation (1), defined in the previous section. Considering a confidence level, z=1.96 for α=0.05, *p* is the estimated proportion = .50, and *e* the desired level of precision = .05, n_o, the sample size, equals 385 organizations. Since the population size is known (*N*=1634), the corrected sample size becomes *n* = 312 organizations by using equation (2). This value marks the end of phase 2 (R3, Figure 2).

3.2 Simple Random Sampling vs. Stratified Random Sampling

Up to now, a suitable sampling frame consistent with the objective was defined, the 1634 organizations of Cuenca city. A representative sample size was also defined: 312 organizations. Based on the main objective, the next phase (phase 3) allows the selection of the most appropriate sampling technique, single random sampling or stratified random sampling. The first is defined by considering all the population as a whole and the second is defined by considering the population divided in sub-sectors.

Single Random Sampling. In this case, all the population defined in the sampling frame was considered as a whole, and the 312 organization were randomly selected by using the Excel tool Data Analysis, creating the organizations list (OL1) (R4, Figure 2).
Stratified Random Sampling. In this case, all the population defined in the sample frame was considered by sub-sector. Each sub-sector is considered a stratum, comprised of cases that should be present in the sample. Table

5 exemplifies this estimation. Each stratum is analysed separately and the same procedure to randomly select the organizations in each one is like the one applied to the whole in the single random sampling, creating the organizations list (OL2) (R5, Figure 2).

Table 5. Sample adjustment by stratum (Source: Data base 3, (SRI, 2017))

ISIC, rev.4 Code (F)	Organizations	%	Sample by stratum (%*n)	Sample adjustment by stratum
F410010	1001	61.26%	191.13	**191**
F432102	186	11.38%	35.52	**35**
F433020	155	9.49%	29.60	**30**
F432101	151	9.24%	28.83	**29**
F439020	141	8.63%	26.92	**27**
Total	**1634**	**100 %**		**n = 312**

3.3 Comparative Analysis

The comparative analysis (Phase 4 of the proposed methodology), consists of answering the question: Are there any differences between the two lists obtained (OL1 and OL2) according to the sampling techniques? The comparison was made by identifying common organizations in both sides, that is, the organizations that were selected by using two sampling techniques (sub section 3.2). After the analysis of the first sampling selection procedure (Tables 6 and 7, part 1) was completed, 55 organizations were identified in both lists (18%): 39 organizations (71%) of the stratum F410010, 5 (9%) of stratum F432102, 2 (4%) of stratum F433020, 2 (4%) of stratum F432101, and 7 (13%) of stratum F439020. The remaining 257 organizations (82%) are not repeated. In this analysis, cases from all stratum were also present in the single random sample, but it may not be necessarily always like that. In the limit, organizations from some of the strata may not be selected.

To validate the sample representativeness, the procedure of sample selection process was repeated four times more (Tables 6 and 7). Tables 8 and 9 summarize the results obtained (average, maximum and minimum) after five sample selection processes.

Table 6. Sample selection process _part 1

Condition	1st time		2nd time		3rd time		4th time		5th time	
	#	%	#	%	#	%	#	%	#	%
Repeated	55	17.63%	63	20.19%	59	18.91%	55	17.63%	56	17.95%
Not Repeated	257	82.37%	249	79.81%	253	81.09%	257	82.37%	256	82.05%
Total	**312**	**100%**	**312**	**100%**	**312**	**100%**	**312**	**100%**	**312**	**100%**

Table 7 Sample selection process in repeated elements by stratum _ part 1

ISIC, rev.4 Code (F)	1st time		2nd time		3rd time		4th time		5th time	
	#	%	#	%	#	%	#	%	#	%
F410010	39	71%	37	59%	34	58%	36	65%	39	70%
F432102	5	9%	5	8%	7	12%	9	16%	2	4%
F433020	2	4%	8	13%	9	15%	3	5%	8	14%
F432101	2	4%	4	6%	4	7%	4	7%	4	7%
F439020	7	13%	9	14%	5	8%	3	5%	3	5%
Repeated	**55**	**100%**	**63**	**100%**	**59**	**100%**	**55**	**100%**	**56**	**100%**

An average of 57.6 organizations were identified in both lists (18.46%): 37 organizations (64.47%) of the stratum F410010, 5.6 (9.77%) of stratum F432102, 6 (10.27%) of stratum F433020, 3.6 (6.24%) of stratum F432101, and 5.4 (14.29 %) of stratum F439020. The remaining 254.4 organizations (81.54%) are not repeated, in average. Somewhat, these results obtained in the sample selection process are in line with Pareto rule 80/20 that say; “80% of the consequences are derived from 20% of the causes”. In other words, only 20% approximately of the data are repeated compared with 80% approximately that are not, relative to the stated research objective (R6, Figure 2). However, and following Sharma (2017) indications, with a stratified random sampling the potential

human bias in the selection of cases is reduced resulting a sample that best represents the population under study.

Table 7. Sample selection process _ part 2

Condition	average		maximum		minimum	
	#	%	#	%	#	%
Repeated	57.6	18.46%	63	20.19%	55	17.63%
Not Repeated	254.4	81.54%	257	82.37%	249	79.81%
Total	**312**	**100%**	**312**	**100%**	**312**	**100%**

Table 8. Sample selection process in repeated elements by stratum _ part 2

ISIC, rev.4 Code (F)	average		maximum		minimum	
	#	%	#	%		
F410010	37	64.47%	39	70.91%	34	57.63%
F432102	5.6	9.77%	9	16.36%	2	3.57%
F433020	6	10.27%	9	15.25%	2	3.64%
F432101	3.6	6.24%	4	7.27%	2	3.64%
F439020	5.4	9.26%	9	14.29%	3	5.36%
Repeated	**57.6**	**100%**	**63**	**100%**	**55**	**100%**

4. Conclusions

There are different ways to determine the sample size of a specific population, but when different characteristics of context are represented, such as country, region, zone, province, city, and sectors with their economic activities, they can influence future inference. According to this situation, an interesting alternative may be to apply a set of criteria related to the research questions in selecting a specific sector and a specific population, and later to select a working sample where the purpose is to be able to make inferences to the population selected.

Thus, an interactive and sequential process to select a sampling method was presented, with four phases and six specific results (see flowchart of the methodology as Figure 2). Cuenca city, in the province of Azuay in southern Ecuador, was selected for the study. There are some important key indicators about the CS in terms of GVA that provided the basis for the criteria to be adopted. The flowchart proceeds with the definition of the CS as the selected sector, and the definition of the population (1634 organizations) that consider 80% of the organizations in Cuenca city, with five categories. Two samples (OL1 and OL2) with 312 organizations were selected with the application of two sampling techniques. A comparative analysis between two lists with five sample selection processes concluded that an average of 18.46% of the organizations were selected in the two approaches. The remaining 81.54% in average of the organizations in both samples are not repetitive.

It is interesting to confront this result with the Pareto rule 80/20. In this work, approximately 20% of the cases are repeated in the sampling process and approximately 80% are not repeated. These two different ways of defining a sample may be related to differential objectives as it regards its future use or application, as emphasized in different works (i.e. Etikan and Bala, 2017; Sharma, 2017; Ramsey and Hewitt, 2005). Both demonstrate a methodology to select a representative sample of a population, having in regard the main objective and the purpose of the study, which may provide some economy of resources in terms of the cases that must be selected, without compromising the sample representativeness.

The four phases and the six results of this methodology could be replicated and adapted to other business and management studies by selecting activities within each phase, adapting them in relation to other contexts and integrating activities. In general, the parameter GVA is used quite often and the quantity of organizations similar to that within the study of the CS is possible to get. So, with this into consideration, the interpretation of the data must be in line with the sampling method used. This methodology presents some limitations in its construction such as the availability and reliability of the data in the web of the Central Bank of Ecuador, the quantity of organizations within the CS defined in the web of the SRI Ecuador and the criteria used to select the data. As such, the main problem can be related to the sources of information, and consequently the results of the sampling technique, and it may also be related to the choice of criteria.

The phases of this methodology can be replicated and adapted to many business and management studies, which face similar circumstances in terms of selecting a representative sample.

5. References

BCEcuador, B. C. (2015) 'Cuentas Nacionales - Valor Agregado Bruto - Provincia - Azuay'. Available at: http://sintesis.bce.ec:8080/BOE/BI/logon/start.do?ivsLogonToken=bceqsappbo01:6400@1847685JyWz2mTJdT0X9hqE2m5wWRz1847683JYE64qNsD00qrmdm72pxZls.

BCEcuador, B. C. (2016) 'Valor Agregado Bruto de las Industrias'. Available at: http://sintesis.bce.ec:8080/BOE/BI/logon/start.do?ivsLogonToken=bceqsappbo01:6400@1847685JyWz2mTJdT0X9hqE2m5wWRz1847683JYE64qNsD00qrmdm72pxZls.

Burnam, M. A. and Koegel, P. (1988) 'Methodology for Obtaining a Representative Sample of Homeless Persons: The Los Angeles Skid Row Study', *Evaluation Review*, 12(2), pp. 117–152. doi: 10.1177/0193841X8801200202.

INEC, I. N. de E. y C. (2017) *Conozcamos Cuenca a través de sus cifras*, *Conozcamos Cuenca a través de sus cifras*. Available at: http://www.ecuadorencifras.gob.ec/conozcamos-cuenca-a-traves-de-sus-cifras/.

INEC, I. N. de E. y C. (2018) *Ecuador en Cifras*, *Instituto Nacional de Estadísticas y Censos INEC (2018) Ecuador en Cifras. Available at: http://www.ecuadorencifras.gob.ec/estadisticas/.* Available at: http://www.ecuadorencifras.gob.ec/estadisticas/.

Israel, G. D. (2008) 'Determining Sample Size 1', (November), pp. 2–7.

Karuppusami, G. and Gandhinathan, R. (2006) 'Pareto analysis of critical success factors of total quality management: A literature review and analysis', *TQM Magazine*, 18(4), pp. 372–385. doi: 10.1108/09544780610671048.

Ramsey, C. A. and Hewitt, A. D. (2005) 'A Methodology for Assessing Sample Representativeness', (October 2004), pp. 71–75. doi: 10.1080/15275920590913877.

Saunders, M., Lewis, P. and Thornhill, A. (2016) *Research Methods for Business Students*. Seventh Ed. Edinburgh Gate: Pearson Education Limited.

SENPLADES, S. N. de P. y D. (2018) *Sistema Nacional de Información*. Available at: http://sni.gob.ec/inicio.

SRI, S. de R. I. (2017) *Registro Único de Contribuyentes (RUC)*, *Diiciembre 2017*. Available at: https://www.sri.gob.ec/web/guest/RUC.

Taherdoost, H. *et al.* (2016) 'Sampling Methods in Research Methodology ; How to Choose a Sampling Technique for', 5(2), pp. 18–27.

United Nations (2008) *International Standard Industrial Classification of All Economic Activities (ISIC), Rev.4*. New York: United Nations Publication. Available at: https://unstats.un.org/unsd/publication/seriesm/seriesm_4rev4e.pdf.

United Nations (2015) *Central Product Classification (CPC) Version 2.1*, *United Nations Statistics Division*. New York. Available at: https://unstats.un.org/unsd/classifications/unsdclassifications/cpcv21.pdf.