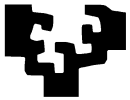
eman ta zabal zazu

universidad del país vasco



euskal herriko unibertsitatea

Department of Financial Economics II

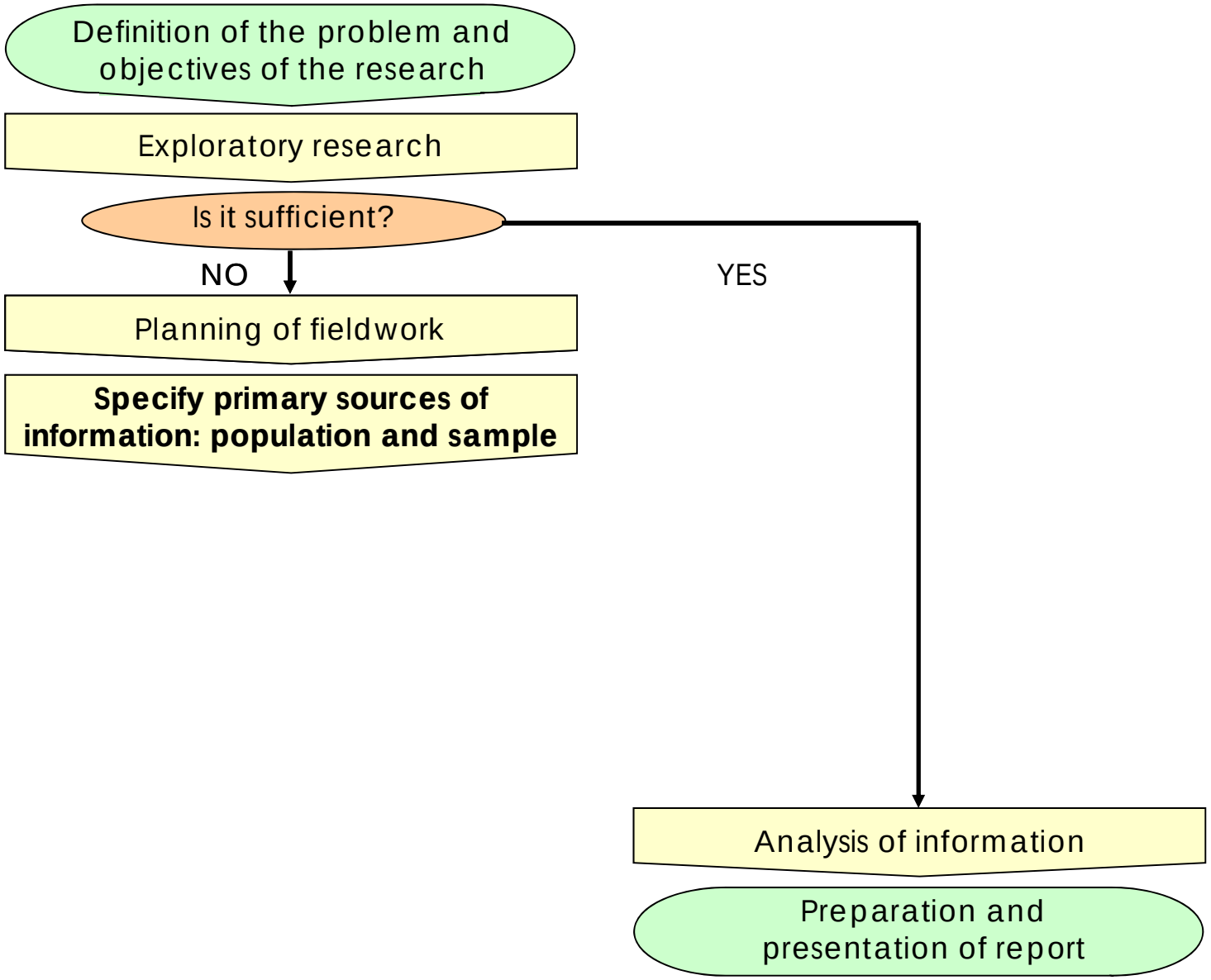
(FINANCIAL ECONOMICS AND ACCOUNTING, MARKETING AND MARKETING RESEARCH)

MARKETING RESEARCH: AN INTRODUCTION

Unit 6: An introduction to sampling

Jon Charterina-Abando

Marketing Research
Sample design



The sampling process

Sampling is the process of selecting a subset that is representative of the population of interest. Only the correct use of the sampling technique will make it possible to extrapolate the results to the whole population.

Marketing Research
Sample design
Census or Sample?

Characteristic	Census	Sample
Population size	*Small*	*Large*
Population variability	*Large*	*Reduced*
Time available	*Plenty*	*Limited*
Definition of population	*Sound*	*Poor or non-existent*
Random error	*Non-existent*	*In principle tolerable and bounded*
Systematic errors	*Greater risk*	*Lesser risk*

Basic concepts

- **Population:** A set of elements, limited to a place and time, from which the information required for the study can be drawn.
- **Element:** Unit of population, usually individuals or events
- **Sample frame:** Representation of the population, usually a list of elements or groups of elements
- **Sample unit:** Element or group of elements, or individual from the frame that represents one or several elements.

Marketing Research
Sample design

Sampling methods:

a).- Probabilistic sampling:

The selection of the sample is made randomly, which predetermines the errors and the precision of the estimates.

b).- Non-probabilistic sampling:

The selection does not allow for the level of random error to be known as it is a non-random selection, left up to the judgement of the researcher and/or the interviewer.

Sampling techniques:

PROBABILITY	NON-PROBABILITY
Characteristics: ▪ Sample units are selected randomly. ▪ The selection probability is known for each element in the population. ▪ Error can be bounded and error precision evaluated because it is based on statistics.	**Characteristics:** ▪ Selection of the sample is not random. ▪ Error precision cannot be calculated and error cannot be bounded, because it is not based on the probability theory.

- Simple random sampling
- Stratified sampling
- Cluster sampling
- Two-stage sampling
- Multi-stage sampling, etc.

- Convenience sampling
- Snowball sampling
- Quota sampling, etc.

Necessary requirements of samples:

1. **All the members of the population must have the same probability to be selected.**

2. **Unless we want it to show up in the results, the selection criteria used must not be related to the responses.**

3. **For probabilistic sampling, procedures to determine a minimum size subject to a given *sampling error* will have to be followed.**

Sampling error:

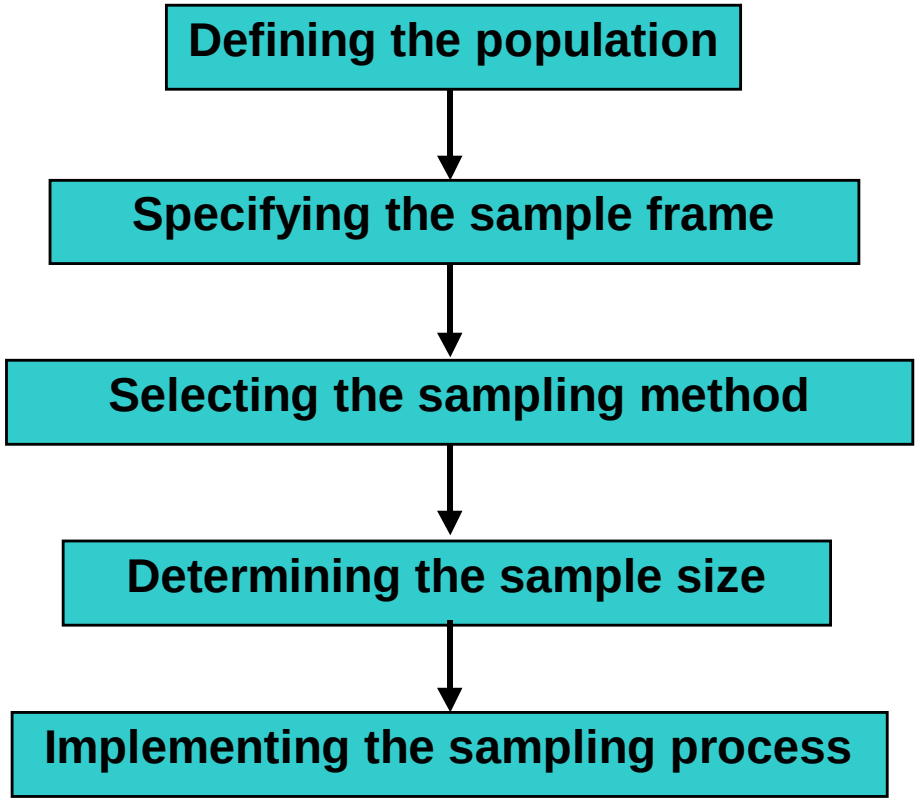
a).- Random or sampling error:

Is the error caused by observing a sample instead of the whole population. In the case of probability sampling, it can be bounded.

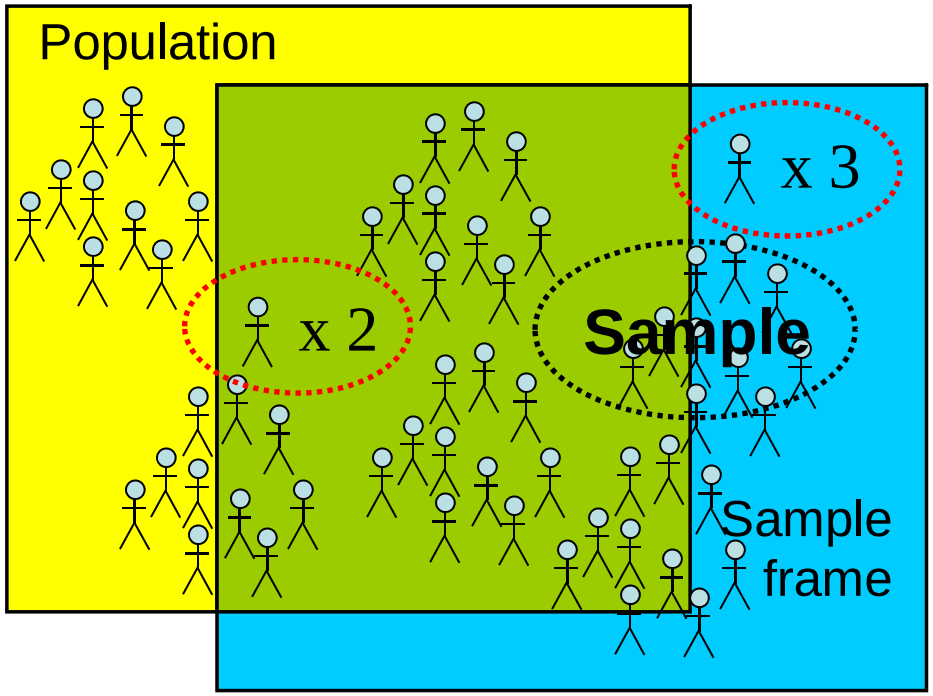
b).- Systematic error:

Is due to errors in the sampling process and cannot be measured. For example, an obsolete frame or list, or a selection of the sample using inappropriate criteria, can result in this type of error.

Stages in the sampling process



Sampling problems in practice



Disparities between the population and the sample frame

Random sample error

<u>Example</u>: Prices corresponding to 40 product reference numbers

P1 = 1,200	P9 = 1,500	P17 = 4,100	P25 = 2,500	P33 = 1,430
P2 = 1,340	P10 = 2,400	P18 = 3,110	P26 = 3,000	P34 = 1,900
P3 = 2,150	P11 = 3,200	P19 = 4,980	P27 = 3,540	P35 = 1,680
P4 = 3,200	P12 = 3,400	P20 = 4,270	P28 = 2,600	P36 = 2,150
P5 = 2,130	P13 = 4,900	P21 = 2,180	P29 = 2,750	P37 = 1,430
P6 = 3,210	P14 = 1,750	P22 = 2,890	P30 = 3,190	P38 = 2,160
P7 = 2,240	P15 = 1,800	P23 = 2,200	P31 = 3,600	P39 = 2,190
P8 = 3,150	P16 = 2,450	P24 = 2,000	P32 = 5,000	P40 = 2,240

Mean: μ = 2,678 € **Standard deviation: σ = 994**

Random sample error

By randomly selecting samples from 5 product reference numbers, we obtain the following (sample) means:

m1 = {P1,P7,P10,P32,P21} = 2,604

m2 = {P11,P31,P12,P3,P17} = 3,290

m3 = {P21,P17,P12,P35,P40} = 2,720

m4 = {P13,P7,P30,P3,P38} = 2,928

m5 = {P16,P5,P1,P29,P24} = 2,106

Random sample error

The above sample means m_i will have a deviation, greater or smaller, from the mean price, taking into consideration all the reference numbers sold in the market (μ):

$$e1 = |\mu - m1| = |2.678 - 2{,}604| = 74$$
$$e2 = |\mu - m2| = |2.678 - 3{,}290| = 612$$
$$e3 = |\mu - m3| = |2.678 - 2{,}720| = 42$$
$$e4 = |\mu - m4| = |2.678 - 2{,}928| = 250$$
$$e5 = |\mu - m5| = |2.678 - 2{,}106| = 572$$

These deviations are called bias or errors.

Random sample error

The bias or error between the mean price of any given sample and the mean price of the population will lie within an interval (called *confidence interval*) bounded by -k and k with a probability of 1 - α or *confidence level*.

$$P\left[-k \leq \frac{m_i - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N}}} \leq k\right] = 1-\alpha$$

Random sample error

Therefore, by finding the value of μ we obtain the confidence interval:

$$P\left[m_i - k\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N}} \le \mu \le m_i + k\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N}} \right] = 1-\alpha$$

Random sample error

Instead of a single value we obtain two values that delimit an interval which contains the true value of μ.

Random sample error

In principle, the more reference numbers from the 5 initial ones we select to calculate the mean of all references, the closer the averages obtained will be to μ (i.e., the value 2,678).

Therefore, we should **know how many observations we must make** in order not to exceed a predefined error level.

Optimal sample size estimators

Parameter	Finite population	Infinite population
Mean (μ)	**Size. Sample mean** $$n_{\bar{x}} = \frac{S^2 N Z_{a/2}^2}{e^2 N + Z_{a/2}^2 S^2}$$	**Size. Sample mean** $$n_0 = \frac{S^2 Z_{a/2}^2}{S^2}$$
Total (T)	**Size. Sample total** $$n_t = \frac{S^2 N^2 Z_{a/2}^2}{e^2 + N Z_{a/2}^2 S^2}$$	
Proportion (P)	**Size. Sample proportion** $$n_p = \frac{pq N Z_{a/2}^2}{e^2 (N-1) + pq Z_{a/2}^2}$$	**Size. Sample proportion** $$n_0 = \frac{p q Z_{a/2}^2}{e^2}$$

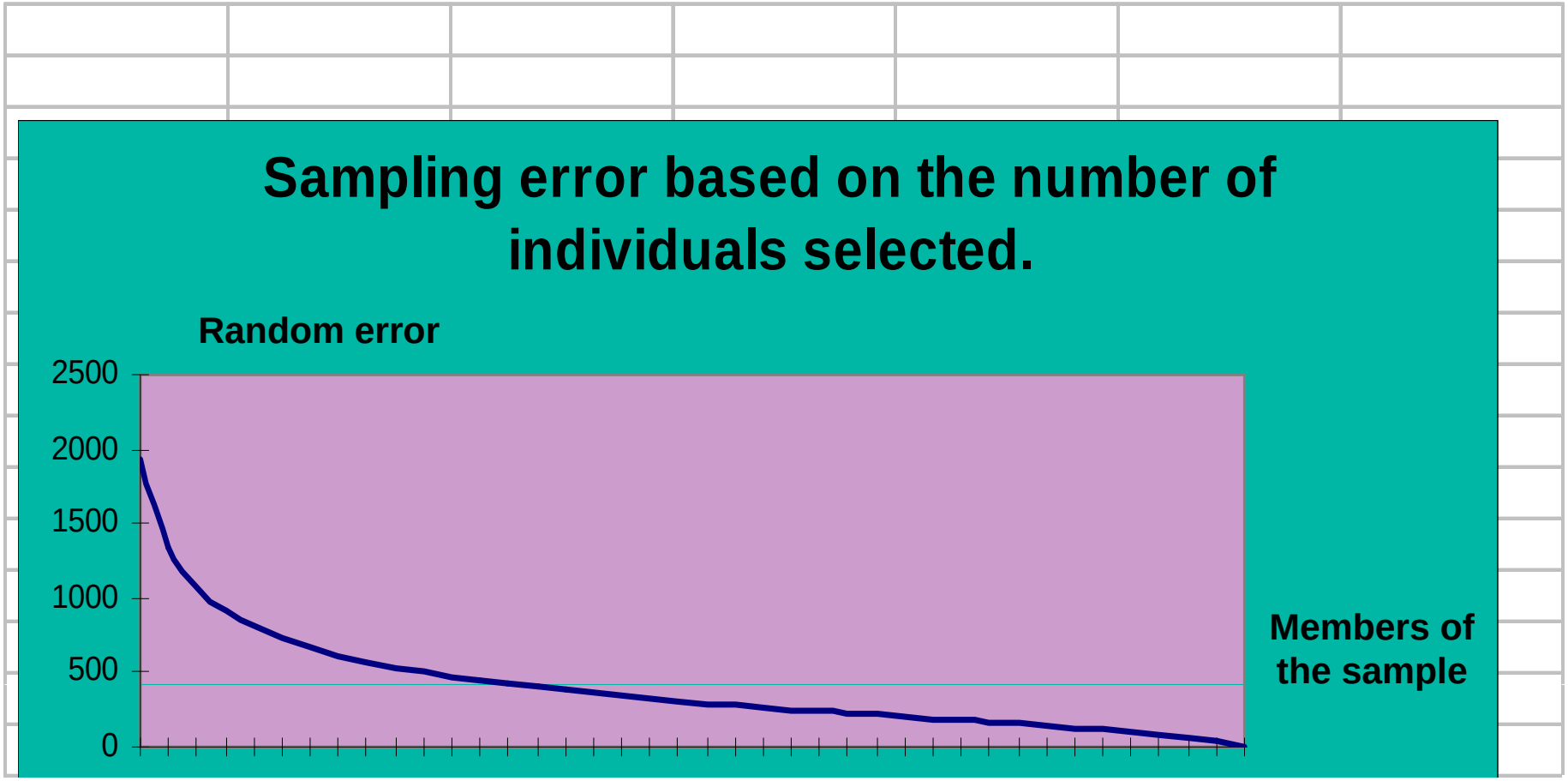
Optimal (minimum) size when the variable to be estimated is a proportion:

$$n_p = \frac{p\,q\,N\,Z_{a/2}^2}{e^2\left(N-1\right)+p\,q\,Z_{a/2}^2}$$

When the population is very large, the following relationship is used:

$$n_0 = \frac{p\,q\,Z_{a/2}^2}{e^2}$$

Random sample error



Sample size correction:

Size n estimated by means of formulas does not take into account non-collected observations. Reasons:

… lack of willingness to collaborate,

…absence,

…other circumstances that may invalidate certain interviews, etc.

Therefore, the initial number of interviews should be increased.

Example: A study of patronage behaviour in Bilbao

Patronage: The relationship of a buyer with an establishment → **OBJECTIVES:**

- To explain why individuals visit or shop at a certain retail establishment or venue, from among a group of possible establishments/venues.
- To measure the relative appeal of each establishment.

Example: A study of patronage behaviour in Bilbao

Sample selection process:

1. The total number of surveys to be conducted is 400, over 11 working days (15 calendar days), an average of approximately 35 surveys per day. → SAMPLE SIZE TABLES

... or at any rate:

$$n = \frac{p.(1-p).k^2}{e^2} = \frac{0{,}5.0{,}5.2^2}{0{,}05^2} = 400$$

Example: A study of patronage behaviour in Bilbao

Sample selection process:

2. The following table shows the fieldwork schedule, including the distribution of surveys throughout the day:

Hours (Mon-Fri)	No. of interviews
9.30 to 13.30h	17
16.30 to 20.30h	18

Example: A study of patronage behaviour in Bilbao

Sample selection process:

3. The interviews will be arranged by age and gender groups, as shown in the following table:

	Male	Female	Total by age
15 to 29 years old	52	47	**99**
30 to 64 years old	111	103	**214**
65 to 74 years old	21	26	**47**
75 years or older	12	28	**40**
Total by gender	**196**	**204**	**400**

Example: A study of patronage behaviour in Bilbao

Sample selection process:

4. Based on the different areas where the interviews will be conducted, there are 3 areas: (1) Bilbao-Centre; (2) Bilbao-Neighbourhoods; and (3) Outskirts-Control Group. The marginal tables for the respective areas are:

(1) Bilbao-Centre and (2) Bilbao-Neighbourhoods (one table for each area):

	Male	Female	Total by age
15 to 29 years old	24	21	**45**
30 to 64 years old	50	47	**97**
65 to 74 years old	9	12	**21**
75 years or older	5	13	**18**
Total by gender	**89**	**93**	**182**

Example: A study of patronage behaviour in Bilbao

Sample selection process:

(3) Outskirts-Control Group:

	Male	Female	Total by age
15 to 29 years old	4	5	**9**
30 to 64 years old	10	8	**18**
65 to 74 years old	2	3	**5**
75 years or older	2	2	**4**
Total by gender	**18**	**18**	**36**

Example: A study of the quality of management in SMEs

Objectives of the project:

- ***To measure the relationship between (total and partial) management quality in SMEs in the Autonomous Community of the Basque Country (ACBC) and the economic results achieved.***
- ***To find out the levels of use and efficiency of certain management tools.***

Some data on the sample frame used for part 1:

DATABASE of DUN & BRADSTREET ESPAÑA, S.A. → 'Duns 5000'

FILE Dunsfile FORMAT EXCEL

No. OF RECORDS 7506

FILE DATA:

DUNS NUMBER

BUSINESS NAME

TELEPHONE NUMBER

FULL NAMES of up to 10 management personnel

Fieldwork conducted for part 1:

Telephone interviews of 943 Basque companies.

Contracted institute: GIZAKER, S.L.

Data sheet:

- ✓Universe or population: 6.825 companies
- ✓943 interviews of managers of Basque companies, based on a hierarchical structure
- ✓Random error e = ± 3.023% with a confidence level of 95.5% (k =2); p=q=0.5, for total data
- ✓Dates interviews were conducted: from 7 to 15 April 2003

POPULATION DISTRIBUTION

		Size according to no. of employees				
		From 10 to 49 employees	From 50 to 99 employees	From 100 to 249 employees	250 employees or more	Total
Primary (SIC 1-14)	Araba	11	2	0	0	13
	Gipuzkoa	27	6	0	0	33
	Bizkaia	61	8	5	0	74
Total Primary Sector		**99**	**16**	**5**	**0**	**120**
Industrial (SIC 15-39)	Araba	537	75	38	20	670
	Gipuzkoa	1065	163	69	52	1349
	Bizkaia	1397	164	113	54	1728
Total Industrial Sector		**2999**	**402**	**220**	**126**	**3747**
Services (SIC 40 or higher)	Araba	366	31	19	8	424
	Gipuzkoa	742	71	41	22	876
	Bizkaia	1377	136	92	53	1658
Total Service Sector		**2485**	**238**	**152**	**83**	**2958**
Total		**5583**	**656**	**377**	**209**	**6825**

THEORETICAL DISTRIBUTION OF SAMPLE

		Size according to no. of employees						
		From 10 to 49 employees	From 50 to 99 empl.	From 100 to 249 empl.	250 empl. or more	Size total	Min. recomm. sample (*)	Sample according to sectors
Primary (SIC 1-14)	Araba	2	2	0	0	4	4	
	Gipuzkoa	4	4	0	0	8	8	
	Bizkaia	8	6	4	0	18	18	
Total Primary Sector		14	12	4	0	30	30	91
Industrial (SIC 15-39)	Araba	33	27	19	13	92	5	
	Gipuzkoa	66	59	34	33	192	11	
	Bizkaia	86	59	56	34	235	14	
Total Industrial Sector		185	145	109	80	519	30	334
Services (SIC 40 or higher)	Araba	23	11	9	5	48	3	
	Gipuzkoa	46	26	20	14	106	6	
	Bizkaia	85	49	45	34	213	12	
Total Service Sector		154	86	74	53	367	30	327
Sample according to size		353	243	187	133	916	90	752
Verification using formulas		345	236	186	133	900		
(Primary S. increment)		8	6	2	0	16		
	Random error	0.05152	0.05094	0.05198	0.05241	0.03075		
	Final random error:	0.04639	0.04807	0.06006	0.07744	0.03023		

DISTRIBUTION OF FINAL SAMPLE ACHIEVED

		Size according to no. of employees				
		From 10 to 49 employees	From 50 to 99 employees	From 100 to 249 employees	250 empl. or more	Size total
Primary (SIC 1-14)	Araba	3	0	0	0	3
	Gipuzkoa	4	4	0	0	8
	Bizkaia	13	5	1	0	19
Total Primary Sector		**20**	**9**	**1**	**0**	**30**
Industrial (SIC 15-39)	Araba	52	21	16	10	99
	Gipuzkoa	67	78	33	22	200
	Bizkaia	109	62	43	21	235
Total Industrial Sector		**228**	**161**	**76**	**43**	**534**
Services (SIC 40 or higher)	Araba	33	12	7	5	57
	Gipuzkoa	47	28	23	11	109
	Bizkaia	101	51	37	24	213
Total Service Sector		**181**	**91**	**67**	**40**	**379**
Sample according to size		**429**	**261**	**160**	**93**	**943**
Random error		**0.04639**	**0.04807**	**0.06006**	**0.07744**	**0.03023**