

SELECTING A REPRESENTATIVE SAMPLE

ABSTRACT

The output of any research work depends, to a reasonable extent, on the adequacy of the sample from which data are obtained for the research. A sample is adequate when it is representative of the population or when it possesses the characteristics that are typical of the population from which it is drawn. The focus of this paper is on how researchers can select samples that are really representatives. Descriptive method was utilized in the writing of the paper. Attempts are made, in the paper, to describe the concepts of sample representativeness, sampling methods and sampling error. Some suggestions are then made on measures to take in carving out truly representative samples.

Keywords: Selection, Sample, Representative sample.

INTRODUCTION

Information Technology has dominated the world today and radical technological change and fusion have changed the way work is organized and performed (Dauda, 2007). To this end, frantic efforts are made to obtain relevant information on different aspects of human life.

Managers are fond of making decisions on different areas of business management in order to achieve organizational objectives. The decision making process requires certain information that will provide essential clues to certain issues. The information so needed could be derived from hundreds or thousands of prospective respondents who could be consumers, suppliers, employees, scholars or government officials. As a result of the largeness of the population being dealt with, a researcher may need to carve out just a fractional segment of the entire population to ensure easy and timely collection of data. Herein lies the relevance of sampling.

It is one thing to create a sample out of a particular population; it is another thing to ensure that the sample is large enough to represent the population. By population, we mean the totality of items in an investigation (Gupta and Gupta, 1987). In other words, population is a collection of the individual items that are to be observed in a given problem situation; the items could be living or non-living things.

A sample is a selected portion or subset of the population being investigated (Oladele, 2007). The elements in the sample possess the same characteristics with the population but the differences are in the size; that is, sample < population (Ayodele et al. 2003).

A quality control manager may decide to take a few cocoa beans, say 50, from a bag of cocoa to investigate the quality level of the content of the bag. The thinking of

40 the quality control manager is that by examining 50 out of the entire cocoa beans in
41 the bag, he will get a fair notion of the quality of the entire population cocoa beans
42 inside the bag.

43 The objective of this paper is to find answers to the following questions:

- 44 i. What are the different methods of creating samples?
- 45 ii. What should be done to ensure selection of a truly representative sample?
- 46 iii. How can sampling errors be reduced?

47 In view of the nature of this study, descriptive method was used in writing, the
48 paper.
49 Towards this end, efforts were made to describe sampling methods, sample
50 representativeness and sampling error.

51

52 **SAMPLING METHODS**

53

54 Kerlinger (1992) defined sampling as the process of taking any part of a
55 population or universe as representative of that population or universe. Another
56 way of saying the foregoing is that sampling is a means of estimating population
57 parameters from only a few items. The process takes place especially when the
58 population being studied is a large one such that it becomes impossible or costly to
59 investigate each item in it (population). What could be done in that situation is to
60 select a few elements from the whole of materials being investigated and then make
61 a generalization about the population (Arora and Arora. 2009). The method by
62 which sample was drawn determines, to a reasonable degree, the extent to which
63 generalizations about the population can be made (Nworgu, 1991). The sampling
64 method so used determines the representativeness of the sample in relation to the
65 population.

66

67 Sampling methods can be categorized into two main classes namely,
68 probability sampling techniques and non-probability sampling techniques.

69

70 **Probability Sampling Methods**

71

72 These are sampling techniques for which we can determine the chance of
73 drawing each member of the population to form a sample. The method describes a
74 situation in which each item within the population has equal chance of being chosen
75 to form a sample (Pearson. 1993). Thus, there is no bias in the selection of sample
76 members. One other major advantage of this set of sampling methods is that it is
77 easy to measure the sampling error and interpret sampling results. As a result of
78 these favourable features, conclusions reached by studying a particular sample are
79 considered generalizable to that population or other similar populations.

80

81 Some of the probability sampling methods are listed below

82 Simple random sampling

83 Stratified sampling

84 Systematic sampling

85 Cluster sampling

86

87 **Simple Random Sampling:** This sampling method Involves selecting a few elements
88 from a total population in such a way that each member of the population has an
89 equal chance of being selected. Thus, a sample drawn at random is unbiased in the
90 sense that no member of the population has any more chance of being selected than
91 any other member (Osuala, 2001).

92

93 One means of carving out a sample from the entire population is to name or number
94 each item in the population. Next, each name or number is cut into a small slip and
95 squeezed and then placed in a container. The investigator puts his hand into the
96 container to pick the number of slips required to obtain the required size of the
97 sample.

98

99 The selection procedure described above may not be possible in a situation where
100 the studied population is very large. For instance, taking a random sample of the
101 population of Lagos State. A table of random numbers, an example of which is
102 shown in Table 1, becomes useful. The computer is often used to accomplish the
103 selection process at this level.

104

Table 1. Random Numbers

105

63271	59986	71744	51102	15141	80714	58683	93108	13554	79945
88547	09896	954336	79115	08303	01041	20030	63754	08459	28364
55957	57243	83865	09911	19761	66535	40102	26646	60147	15702
46276	87453	44790	67122	45573	54358	21625	16999	13385	22782
55363	07449	34826	15290	76616	67194	18277	21151	68684	08263
69393	92785	49902	58477	42048	30378	87618	26933	40640	16281
13186	29431	88130	04588	38733	81290	89541	70290	40113	08243
17726	28652	56836	78351	47327	18518	92222	55201	27340	10493
36520	64465	05550	30157	82242	29520	69753	72602	23756	54935
81628	36100	39254	56835	37636	02421	98063	89641	64953	99337

84694	48968	75215	75498	49539	74240	03466	49292	36401	45525
63231	11618	12631	75055	43915	26488	41116	64551	56827	30825
70502	53225	03655	05915	37140	57051	28393	91322	25653	06543
06426	24771	59935	49801	11082	66762	94477	02494	88215	27191
20711	55609	29430	70165	45406	78484	31639	52009	18873	96927
41990	70538	77191	25860	55204	73417	83920	69468	74972	38712
72452	36618	76298	26678	89334	33938	95567	29380	75906	91807
37042	40318	57099	10528	09925	89773	41335	96244	29002	46453
53766	52875	15987	46962	64342	77592	57651	95508	80033	69828
90585	58955	53122	16025	84299	53310	67380	84249	25348	04332
32001	96293	37203	64516	51530	37069	40261	61374	05815	06714
62606	64324	46354	72157	67248	20135	49804	09926	64419	29457
10078	28073	85298	50324	14500	15562	64165	06125	71353	77669
91561	46145	24177	15294	10061	98124	75732	00815	83452	97355
13091	98112	53959	76607	52244	63303	10413	63839	74762	50289

106 **Source: Anderson, David R. et al (2003), Pg. 258.**

107 **Stratified Sampling:** According to Anderson et al (2003), stratified sampling method
108 is best suited to populations that have different sets of groups within them. In other
109 words, the sampling method is mostly used when dealing with heterogeneous
110 populations. For instance, if a researcher wants to collect relevant data on a topic that
111 says “Life after death” from a set of people, the best sampling method to use is
112 stratified sampling. The reason being that the population being sampled will
113 comprise people of different religious beliefs who are bound to have different
114 opinions on the subject matter. In this case, the heterogeneous population will have
115 to be divided into three homogeneous groups or strata as follows; Christians,
116 Muslims and Traditional believers.

117

118 Nworgu (1991) maintained that there are two types of stratified sampling
119 namely, proportionate stratified random sampling (**PSRS**) and disproportionate
120 stratified random sampling (**DSRS**). In **PSRS**, the population is first stratified in
121 terms of one or more variables of interest to the researcher. Elements are drawn
122 randomly from each stratum in such a way that the relative proportions of the strata
123 in the resultant sample are the same as exist in the parent population. This is saying
124 that the relative contribution of each stratum in the population is exactly its relative
125 contribution in the sample.

126

127 Nworgu (1991) and Oladele (2007) were of the view that **PSRS** ensures
128 greater representativeness of the sample relative to the population and guarantees
129 that minority constituents of the population are represented in the sample. Table 2

130 below illustrates **PSRS** with a population of 1220 entrepreneurs.

131

132 Table 2: Distribution of a population of 1220 entrepreneurs according to
133 categories/sizes.

134

	Entrepreneurs			
	Small-Scale	Medium-Scale	Large-Scale	Total
Population Size	549	427	244	1220
Proportion	0.45	0.35	0.20	1.00

135

136 Table 3: Proportionate stratified random sample for a population of 1220
137 entrepreneur.

138

	Entrepreneurs			
	Small-Scale	Medium-Scale	Large-Scale	Total
Population Size	549	427	244	1220
Proportion	0.45	0.35	0.20	1.00
Sample Size	110	85	49	244

139 In case of disproportionate stratified random sampling, the relative proportions of
140 the strata in the sample do not correspond to their relative proportions in the
141 population. Some strata may be under-represented or over-represented in the
142 sample. Obviously, this sampling mode allows the researcher the freedom of
143 weighting the various strata in any manner he considers fit. Though the method
144 does not make for proper representativeness, Oladele (2007) was of the opinion that
145 the **DSRS** method is preferred where the researcher believes that there is likely to be
146 great within-stratum variation in responses or if he has a particular interest in one or
147 more strata.

148

149 **Systematic Sampling:** It is also called quasi-random sampling. In this method, the
150 first sample element is randomly chosen from numbers 1 through K and subsequent
151 elements are chosen at every Kth interval (Kerlinger, 1992). K will be determined by
152 the size of the sample required. For example, if the population of Texy University is
153 16,000 and a list of all the students making up the population is available. If a
154 sample of 200 students is to be taken, the selection of every 80th student will give the
155 required sample.

156

157

The value of K above is determined as follows.

158

159

$$\frac{\text{Total Population}}{\text{Required sample}} \quad \text{That is, } \frac{N}{n}$$

160

161

162

Therefore,

163

164
$$K = \frac{16,000}{200}$$

165
$$= 80.$$

167 It must be noted that systematic sampling is used only when the list of all items in a
168 population is available as in the case of a class register. The question as to how to
169 determine the first element that will form the sample is answered by choosing a
170 number at random between 1 and 50.

171
172 **Cluster Sampling:** Cluster sampling is otherwise known as area sampling. It is
173 successive random sampling **units** or sets and subsets (**Arikpo, M.** 1986). In selecting
174 a sample using this method, the population (or geographical area) is divided into
175 units or segments with well-spelt-out boundaries. A specified number of these units
176 or a section is drawn. All elements in the units or sections drawn now constitute the
177 sample.

178
179 As observed by Oladele (2007), cluster sampling is used when it can be
180 recognized that some populations are distributed in clusters or groups of settlement
181 and these clusters are to be used as the basis for sampling.

182
183 Ferguson (1981) affirmed that unlike stratified sampling, using cluster
184 sampling does not require a list of the elements in the population before sample can
185 be drawn. As long as these are distinguished clusters or geographical locations,
186 creating samples becomes feasible.

187
188 **Non-Probability Sampling Methods**
189 Non-probability sampling is non-scientific approach to sample formation (Monga,
190 2000). It is the process of getting samples from populations without following any
191 statistical rules. The researcher or investigator uses only his intuition to select sample
192 members is fraught with bias and partiality simply because each item in the
193 population does not have an equal chance of being selected. The ultimate
194 consequence of this abnormality is that it becomes difficult to measure the sampling
195 error and to interpret results.

196 The following techniques are examples of non-probability sampling methods:

- 197 Judgmental sampling
- 198 Quota sampling
- 199 Convenience sampling
- 200 Snowball sampling

201 **Judgmental Sampling:** This involves the use of the researcher's reasoning and
202 judgment to obtain a sample, what determines whether an item in the population
203 will be selected or not into the sample is the personal preference of the investigator.

204 Sampling here is influenced by the personal bias of the person or group of persons
205 selecting members for the sample. That is not to say, however, that judgment
206 sampling is a complete write-off. It is advantageous in the sense that it saves time as
207 the process does not require any listing or numbering of the population or random
208 number tables (Osuala, 2001).

209 **Quota Sampling:** According to Nworgu (1991), this involves selecting those
210 elements that have specific characteristics of interest to the researcher and are
211 accessible to him. This type of sampling is used to ensure that specific elements will
212 be included.

213 Evidently, quota sampling gives room for the researcher or investigator to include
214 any category of the population that is of particular interest to him. It is quicker,
215 easier and cheaper. It is disadvantageous in the sense that the resultant sample is
216 highly biased and thus, cannot be said to be representative (Gupta and Gupta, 2007).

217 **Convenience Sampling:** The major consideration of the researcher using this
218 sampling method is the ease of selecting the sample. The researcher makes no
219 attempt to bring in any element of randomness. He selects members of the sample in
220 a way that is convenient or easy for him.

221 As an illustration, a student of Adekunle Ajasin University, Akungba-Akoko,
222 Nigeria, researching on the topic "The problem of small-scale business entrepreneurs
223 in Ondo State" may decide to get the required sample of small-scale business
224 entrepreneurs from Ikare-Akoko and Owo. This is because the two towns are very
225 close to his institution of research and so, can easily or conveniently collect the
226 needed data from respondents in the two towns.

227 In fact, the researcher does not know how well a Convenient Sample will represent
228 Sample the population regarding the traits or mechanism under research. What
229 makes convenience sample so unpredictable is their vulnerability to severe hidden
230 biases. Leineer, D.J (2014).

231 Therefore, in convenience sampling, the individuals selected by the researcher may
232 not be applicable to the research problem.

233 Hence, there is a rise of collecting poor quality data due to poor outcomes such as,
234 difficulty to convince others to accept the findings or research based on poor
235 foundation. Opong, S.H (2013).

236 **Snowball Sampling:** This method is often used to obtain samples in situations
237 where there is no adequate list which could be used as a sampling frame (Osuala,
238 2001). Towards this end, the researcher contacts a member of the population of
239 interest or identifies a group of respondents who possess the traits desired for the
240 research work (Oladele, 2007). This set of people will in turn, identify another set of
241 people suitable for the research work. This chain continues.

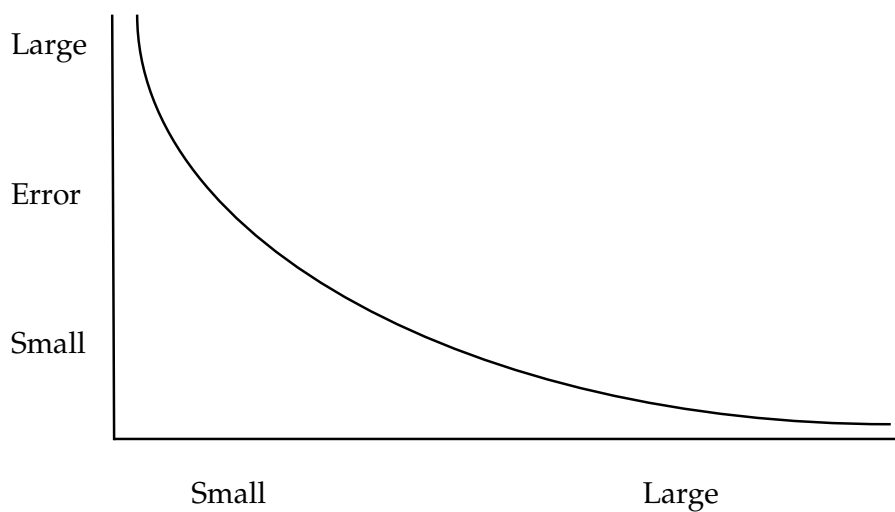
242

243 **SAMPLING ERROR**

244 Sampling error simply means deviation from population values. It is the difference
245 between the result obtained from a sample and the result which would have been
246 obtained from the population. This type of error often occurs when tile complete
247 survey of the population is not carried out, but a sample is taken for estimating the
248 characteristics of the population (Green Tull, 1978).

249 When the entire population is considered no sampling error occurs. The occurrence
250 of sampling error when a sample is carved out of a population depends on the size
251 of the sample relative to the population. As shown in Figure 1, the smaller the
252 sample the larger the error, and the larger the sample the smaller the error.

253
254
255
256
257
258
259
260
261
262
263
264
265
266



267
268
269

270 Size of Sample
271 Figure 1: Sample size vis-à-vis sampling error

272
273 It is possible to measure the sampling error (Pearson. 1993). The error is
274 measured by the standard error of the statistics in terms of probability under the
275 normal curve. The result of the measurement indicates the precision of the estimates
276 of the population based on the sample study (Osuala, 2001).

277
278 Sampling error is derived by calculating the standard error of sample mean. It is the
279 standard deviation of the sampling distribution of sample means (Arora and Arora,
280 2009). It is denoted by σ_x and is given by

281
$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

282

283 Where
284 σ = Standard deviation of the population
285 n = sample size

286

287 It is often advocated that large samples should be used. They are advocated
288 in order to give the principle of randomization a chance to work (Kerlinger, 1992).

289

290 **SAMPLE AND REPRESENTATIVENESS**

291 The word “representative” means to be typical of a population; that is, to
292 exemplify the characteristics of the population. The fact that a segment of a
293 population is taken as representative of that population does not mean that the sample
294 so taken is representative. For example, a researcher investigating a banking habit in
295 First Bank Plc, may decide to draw his sample from two branches of First Bank in
296 Lagos, Nigeria, thereby assuming that the two branches represent the total
297 population of the bank's branches. That sample selection may be wrong as data and
298 opinions collected from the two branches may not be representative of what actually
299 happens in the entire First Bank Plc network. That goes to show that too small
300 samples are not good enough as they tend not to represent the characteristics of the
301 population. This will eventually result in getting results that are likely to be lacking
302 in validity.

303 Kerlinger (1992) opined that in research, a “representative sample” means
304 that the sample has approximately the characteristics of the population relevant to
305 the research in question. If sex and socio-economic class are characteristics relevant
306 to the research, a representative sample will have approximately the same
307 proportions of men and women and middle class and working class individuals as
308 the population.

309

310 The question as to how large a sample must be to be adequate or to be truly
311 representative is not a simple one. Each situation presents its own problems. If the
312 phenomena under study are homogenous, a small sample is sufficient. If units under
313 study are variable, a much larger sample is needed. The greater the variability of the
314 phenomena, the greater the difficulty of obtaining an adequate sample. But that is
315 not to say that using very large samples is always wise. Making use of samples that
316 are too large amounts to waste resources - money, time and energy.

317

318 In order to obtain a representative sample, the following points need to be
319 borne in mind by the researcher.

320

- 321 1. There is need for care and precision on the part of the researcher. He must
322 take care to see that the sample drawn from the population is not biased.
323 Towards this end, it is preferable to use probability sampling methods in
324 selecting sample members since such a procedure guarantees that every
325 population element has equal chance of being selected.
- 326 2. The population being investigated must be properly defined. It should be
327 defined in terms of four things namely, the element, sample unit, extent and
328 time (Moser and Kalton, 1979). For instance, a survey of consumers might
329 specify a relevant population as follows:
- 330 3.

331	Elements	:	Male
332	Sampling Unit	:	27-35 years
333	Extent	:	Kaduna State, Nigeria
334	Time	:	As at 30 th October, 2016

335

336 In the alternative, the population for a study designed to measure buyers' reaction to
 337 a new pharmaceutical item may be the following:

338	Element	:	Pharmacists
339	Sampling Unit	:	Pharmaceutical companies buying over
340			N 5 million worth of item per year.
341	Extent	:	Enugu State, Nigeria
342	Time	:	Year 2014

343 The essence of defining studied population properly is to ensure that whatever data
 344 are obtained from the selected sample meet the researcher's expectation in terms of
 345 data currency, respondents/interviewer's level of experience and
 346 geographical/professional coverage.

347

- 348 4. The usual thing to do after identifying a population is to obtain a complete
 349 and current list of all elements in the population. Nonetheless, in order to get
 350 a representative sample, the researcher should avoid drawing a sample by
 351 preparing an alphabetical list and proceeding down the list until he has
 352 included a sufficient number in the sample. Some letters of the alphabet
 353 include more names from certain groups than others, and this may produce
 354 bias in the sample on the basis of name alone.
- 355 5. The idea of allowing interviewers to select interviewees who are like
 356 themselves should be discouraged. Such a habit results in a biased sample.
 357 For example, if interviewers or researchers are members of a particular
 358 religion, they will likely select many interviewees from the same religion and
 359 include just a few individuals from other religions. Samples obtained through
 360 such an awkward procedure will produce opinions and data that are not
 361 representative of the entire population (Gupta and Gupta, 2007).
- 362 6. Oladele (2007) in his own contribution opined that samples must be adequate
 363 enough to generate generalizable result of findings. There is need for
 364 researchers or investigators to ensure that the elements that make up a
 365 particular sample are those that possess the features that are typical of the
 366 entire population.

367

368 CONCLUSION

369 The validity of the outcome of any research depends, to some extent, on the
 370 extent of adequacy of the sample used in the course of the research. Monga (2000)
 371 stated that the size of sampling error indicates the reliability or precision of the
 372 research. According to him, as the sample size increases, the error decreases and
 373 hence larger samples are considered more reliable and more representative than
 374 smaller samples.

375 In view of the fact that representative samples invigorate and accelerate the

376 process of research, there is need for researchers' training in the special area of
377 "sample representativeness", According to Bushnell (1990), such training could lead
378 to improved research and accelerated technological innovation.

379 In this paper, some suggestions are made regarding how representative
380 samples can be carved out from designated universe or populations. If the
381 suggestions are strongly adhered to by researchers, random fluctuations or sampling
382 errors will be minimized. The lower the sample errors, the more representative a
383 sample becomes.

384

385 REFERENCES

- 386 Anderson, David R; Sweeney, Dennis J & Williams, Thomas A, (2003). *Modern*
387 *Business*
388 *Statistics*, Australia: South Western.
- 389 Arikpo, M. (1986). *Foundation of Social Science Research: A Methodological Guide for*
390 *Students*. Enugu: ABIC Publishers.
- 391 Arora, P.N. & Arora. S. (2009). *Statistics for Management (4th ed)*. New Delhi: S. Chand
392 *and Company Ltd*.
- 393 Ayodele, Sam O., Adegbile Joseph A. & Adewale J. Gbenga (2003). *Evaluation Studies*,
394 Ibadan: Powerhouse Press and Publishers.
- 395 Bushnell. D.S. (1990), Input, Process, Output: A model or Evaluating Training.
396 *Training*
397 *and Development Journal*, London, March, p.31.
- 398 Cresswell, J.W, and Plano Clerk, V.L (2011). Designing and conducting mixed
399 method
400 research (2nd ed.) Thous and Oaks, CA; Sage.
- 401 Dauda, Adeleke (2007), Research and Development for Organizational
402 Competitiveness. *Management Discoveries*,1 (1), 16-18.
- 403 Ferguson, G.A. (1981), *Statistical Analysis in Psychology and Education*(4th ed). New
404 York:
405 McGraw-Hill.
- 406 Green, Paul E. & Tull, Donald S. (1978). *Research for Marketing Decisions*(4th ed). New
407 Jersey: Prentice-Hall, Inc.
- 408 Gupta, C.B. & Gupta, Vijay (1987). *An Introduction to Statistical Methods*. New Delhi:
409 Vikas Publishing House PVT Limited.

410 Gupta, S P & Gupta. Archana (2007). *Statistical Methods*. New Delhi: Sultan Chand
411 and
412 Sons.

413 Kerlinger, Fred N, (1992). *Foundation of Behavioural Research* (3rd ed). New York:
414 Harcourt College of Publishers.

415 Lawrence A Palinkas, Carla A Green, Jennifer P Wisdom, and Kimberly Eaton
416 Hoagwood (2013). Purposeful Sampling for Qualitative Data Collection and
417 Analysis in Mixed Method implementation Research. Research Gate.

418 Leineer, D.J (2014). Convenience Samples and Respondent Pools. 1-36.

419 Monga, G.S (2000). *Mathematics and Statistics for Economics* (2nd ed). New Delhi:
420 Vikas Publishing House PVT Ltd.

421 Moser, A. Claus & Kalton, Graham (1979). *Survey Method in social Investigation* (2nd
422 ed).
423 London: Heinemann Educational Books Ltd.

424 Nworgu, B.G. (1991). *Educational Research: Basic Issues and Methodology*. Ibadan:
425 Wisdom Publishers Limited.

426 Oladele, Olajide (2007). *Introduction to Research Methodology*. Lagos: Niyak Print and
427 Publications.

428 Oppong, S.H. (2013). The problem of Sampling in Qualitative research. Asian Journal
429 of
430 Management Services and education, 1-9.

431 Osuala, E.C. (2001). *Introduction to Research Methodology* (3rd ed). Onitsha:
432 Africana
433 Publishers Limited.

434 Pearson, W.A. (1993). *Management Development for Scientists and Engineers*.
435 Research
436 *And Technology Management Journal, January/February, 19-21.*

437 Walliman, N. (2011). *Research Methods: The basics*. New York: Routledge.

438 Zuli; H.L (2014). A Comparison of Convenience Sampling and Purposive Sampling.
439 Pub Med, 105-11