International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 9, September - 2022, Pages: 88-92

How to Grab And Determine the Size of The Sample For Research

Irfan Ananda Ismail[1]*,Niza Lian Pernadi[2], Agnes Febriyanti[3],

Universitas Negeri Padang, Master of Chemistry Education Study Program
Padang, Indonesia
halo@irfanananda28.com

Abstract: This study aims (1) to find out how to randomly select a sample with various types of cases that exist in the social and community environment influences of the family environment, peers, and perceptions of a profession. Sampling is a process of several objects or populations that will be examined for their properties. The statistical method related to drawing conclusions through samples, is called inductive statistics / inference. Instead, the conclusion of a population or group that does not represent a larger group, is called descriptive / deductive statistics. and the way the data is collected is called a census. The selection of samples is a very important part of the study, measurement is often entrusted to the interview and the list of questions used. In non-experimental studies, researchers were only able to regulate general variables through stratification / grouping of respondents, selection of respondents as samples, making a list of questions and training interviewers. However, even though there are good samples, the question dimensions tested, trained interviewers, researchers cannot supervise and manage the relationship between the respondent and the interviewer, so the results of the information collected can deviate from what was expected. Therefore, the interviewer must really live up to the purpose and objectives of the research, be interested and also want to know the answers to the questions asked. The selection of appropriate / good respondents, it is very necessary to guarantee that the information obtained really reflects the desired population. There are several ways of collecting samples, namely simple random sampling (Simple random sampling), stratified random sampling, systematic sampling, group (cluster sampling), gradual (Multistage sampling) and selected sampling. As for how to take and process samples with nominal scale techniques, ordinal scale, interval scale and constellation scale. Calculations can also be done with various formulas such as cross-sectional research sample formulas and case control and cohort research sample formulas.

Keywords—*component; retrieval, sample, website*

1. Introduction

Sampling is very important for a study, because errors in sampling will lead to incorrect conclusions about an object or thing (population) that we want to study. Therefore, good planning in sampling (sampling design) is very necessary Sam pel retrieval is a process of selecting several objects or populations to be studied for their properties. The sample we take is part of that and population. In other words, a representative hams sample, in the sense that the traits of the population should be reflected through the sample.

The statistical method related to drawing conclusions through samples, is called inductive statistics / inference. Instead, the conclusion regarding a population or group that does not represent a larger group, is called descriptive/deductive statistics. and the way the data is collected is called a census. Data collection through censuses is not always possible given the infinite or very large population, so inference statistics are of great significance in research. The use of inference statistics in research, is considered on the basis of several reasons, including:1. Research can be carried out quickly and cheaply.2. Research can produce more complete and well-encompassing information.3. Since the number of samples is small, the implementation can be carried out more thoroughly.

2. Literature Review

Sample selection is a very important part of the research process. In research, measurements are often entrusted to interviews and lists of questions used. It was actually the respondent, and not the interviewer who made the measurement based on the respondent's own experience. In non-experimental studies, researchers were only able to regulate general variables through stratification / grouping of respondents, selection of respondents as samples, making a list of questions and training interviewers. The success of non-experimental research lies mainly in its success in all the activities mentioned above. However, even though there are good samples, the question dimensions tested, trained interviewers, researchers cannot supervise and manage the relationship between the respondent and the interviewer, so the results of the information collected can deviate from what was expected. Therefore, the interviewer must really live up to the purpose and objectives of the research, be interested and also want to know the answers to the questions asked. More important than all that is that the interviewer must be honest, responsible, clear in writing and observant(Bhisma-Murti, 1997).

The selection of appropriate / good respondents, it is very necessary to guarantee that the information obtained really reflects the desired population. Once the respondent has been selected and the interview process has been completed, then errors in the sampling of respondents or samples, can generate

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 9, September - 2022, Pages: 88-92

data (information), which does not describe the state of the population under study

3. Research Methods

There are several ways of *collecting* samples, namely *simple random sampling (*Simple random sampling), stratified random sampling, *systematic sampling*, group *(*cluster sampling), *gradual* (Multistage sampling) and *selected sampling*(Suparmoko, 1987). Random sampling of mummies provides an equal opportunity for every member in the population, to be selected as a sample. Random strata are used for populations consisting of different groups or strata. Strata or groups can be determined geographically, the level of income. conditions of the territory, employment, education etc. Each group was sampled proportionally to the overall number of samples(GW & Cochran WG, 1967). Furthermore, illustrations will be given for each of them method to be clearer.

a. Simple random sampling

For example, the population to be studied is SLA class III students (sampling unit), in Johar Baru village, Central Jakarta. Next, we compile a sampling framework (Frame), which is a list of names of SLA class HI students in the Johar Baru sub-district area (let's say a number of 500 students), given a sequence number from no.l to 500. Furthermore, a number of samples are selected (for example 100 students) randomly using a list of random numbers or generated numbers 1 to 500, then 100 numbers are taken in a lottery from the 500 student numbers mentioned above (this method is less practical for large sampling)(Anggarini, n.d.).

b. Stratified Random Sampling

Of those 500 students in the example above, students are grouped according to the type of education, for example consisting of 300 high school students, 100 smea people and 100 STM. Then the total sample of 100 people must consist of 60 high school students, 20 smea people and 20 STM students.

c. Systematic sampling

Can be done as follows. From the example above, 500 students will be sampled 100 students. This means that every 5 members of the population (500/100) 1 person is taken as a sample. To determine the first sample selected as a sample, by randomly selecting (draw) one number from the first five numbers, is no. 1,2,3,4 and 5 (suppose the selected one is no.3). Furthermore, the number chosen is the next five-number number, namely the number 8; 13; 18; 23; Etc.

d. Cluster Sampling

Sampling on a cluster/group/swarm basis, is almost the same as stratified random sampling. In the sampling cluster each sub-group consists of a heterogeneous population, while the strata in stratified random sampling subgroup consist of a homogeneous population. For example, cluster sampling of the population of 500 SLA students above, grouped according to Public and Private SLAs (e.g. coming from State SLA 200 students and Private SLA 300 students). So the number of samples taken was 40 State SLA students and 60 Private SLA students.

e. Multistage sampling

Still using the example above, sampling is gradually acknowledged as follows. Of the 500 SLA Class III students above, let's say they come from 10 SLAs in the Johar Baru sub-district area. The first stage is to give a sequence number from no.l to 10 of all SLAs. Furthermore, with consideration of cost, energy and time, randomly selected 4 SLAs from the 10 SLAs above. The second stage is to select 100 samples of students from the 4 selected SLAs. So from the 4 SLAs randomly selected / systematically 25 grade III students each.

f. Selected sampling

This selection of samples essentially ignores the principles of probability and the result is only a rough picture of a situation. This method is done when faced with a situation where the cost is very small, the results are needed immediately and do not require high accuracy, because it is just a general idea. The method is only to determine the properties of the indifidu that need to be taken, the multiplicity and selection of individuals is up to the researcher.

g. Probability Proportionate to Size

Sampling in this PPS way is a variation of stratified sampling with large PSUs that are carried out proportionally

4. Results and Discussion

1. Slovin Method

The Slovin formula uses a normal distribution approach, p=0.5, with the limit value of the error can be determined by the researcher. A reconciliation formulated by Slovin (Steph Ellen, eHow Blog, 2010; with reference to Principles and Methods of Research; (Supranto, 2000) as follows.

- n = N/(1 + Ne^2)
- n = Number of samples
- N = Total population (total number of members of the population)
- e = Error tolerance (tolerance of error; degree of significance; for social and educational is usually 0.05) –> (^2 = rank two). If the population is 2000,

assuming an accuracy rate of 95%, then an error of 5%(0.05) then :
N = 1000,
Significance Level = 5%
so:
n = N/(1 + Ne^2) = 2000/(1 + 2000 x 0.05 x 0.05)
= **333 people.**

2. Krejcie And Morgan's Method
Krejcie and Morgan's method uses the chi-quadrat approach, p=0.5, with an assumed error limit of 5% (0.05) Krejcie and Morgan (1970) in (Sujana, 1972)making a list that can be used to determine the number of samples

Calculating The Size Of The Research Sample

In inferential statistics, the size of the sample largely determines the representation of the sample taken in describing the study population. Therefore, it becomes a need for each researcher to understand the correct rules in determining the minimum sample in a study. How to calculate the sample size of a study is largely determined by the research design used and the data taken. The type of observational research using cross-sectional design will be different from case-control study and khohor, similarly if the data collected is proportional it will be different from if the data used is continued data. In research in the field of public health, most use design or cross-sectional or cross-sectional approaches, although there are some who use case control or khohor. There are many formulas for calculating the minimum sample size of a study, but in this article will be presented a number of formulas that are most often used by researchers

1. **Cross-sectional Research Sample Formula**
For survey research, usually the formula that can be used uses binomunal proportions (binomunal proportions). If the size of the population (N) is known, then it is searched using the following formula:

$$n = \frac{Z^2_{1-\alpha/2}\, p\,(1-p)\, N}{d^2(N-1) + Z^2_{1-\alpha/2}\, p\,(1-p)}$$

Cross Sectional Sample Formula

With a known population (N), researchers can randomly take samples). However, if the population size (N) is unknown or (N-n)/(N-1)=1

then the sample size is calculated by the following formula :

$$n = \frac{Z^2_{\alpha}\, p\, q}{d^2} = \frac{Z^2\ p\,(1-p)}{d^2}$$ (Snedecor GW & Cochra (Lemeshowb dkk, 1997)

Lemeshow Formula(Lemeshow & David W.H.Jr, 1997)

Information:

- n = minimum number of samples required
- = degree of trust
- p = proportion of exclusively breastfed children
- q = 1-p (proportion of children who are not exclusively breastfed
- d = limit of error or absolute precision
  If set =0.05 or Z1- /2 = 1.96 or Z2

1- /2 = 1.962 or rounded to 4, then the formula for the known magnitude N is sometimes changed to:

$$n = \frac{4\ p\, q}{d^2}$$

Simplification of the Lemeshow Formula

For example, we want to find a minimal sample for a study looking for determinants of exclusive breastfeeding. To get a p-value, we have to look at the existing research or literature. From the results of Suyatno's (2001) study in the Demak-Central Java area, the proportion of babies (p) who were given exclusive breastfeeding food was around 17.2%. This means that the value of p = 0.172 and the value of q = 1 – p. With limits from error (d) is set to 0.05 and *the value of Alpha* = 0.05, then the number of samples required is:

$$n = \frac{1{,}96^2\ .\ 0{,}172\ .\ 0{,}828}{0{,}05^2}$$

= 219 people (minimum number)
If no p value is found from other studies or literature, then a maximum estimation of p = 0.5 can be carried out. If you want to be meticulous then the value of d is about 2.5 % (0.025) or smaller.

Deep dive with Slovin Formula

1. **Case Control and Cohort Research Sample Formula**

The formula used to find the sample size of both the case control and the cohort is the same, especially if it uses a proportional measure. It's just that for cohort research, there are also those who use the continuous data size (mean value). The sample size for case control research is aimed at finding a minimal sample for each case group and control group. Sometimes researchers make comparisons between the number of sample groups of cases and controls not necessarily 1 : 1, but also bisa 1 : 2 or 1 : 3 with the aim of obtaining better results. The formula that is widely used to find a minimum sample of case-control research is as follows:

$$n = \frac{(p0.q0 + p1.q1)(Z_{1-\alpha/2} + Z_{1-\beta})^2}{(p1 - p0)^2}$$

Keterangan :

- n = jumlah sampel minimal kelompok kasus dan kontrol
- $Z_{1-\alpha/2}$ = nilai pada distribusi normal standar yang sama de kemaknaan α (untuk α = 0,05 adalah 1,96)
- $Z_{1-\beta}$ = nilai pada distribusi normal standar yang sama d (*power*) sebesar diinginkan (untuk ß=0,10 adalah 1,28
- p0 = proporsi paparan pada kelompok kontrol atau tidak sak
- p1 = proporsi paparan pada kelompok kasus (sakit)
- qo = 1 – p0 dan q1 = 1 – p1

Case Control and Cohort Sample Formulas

In the study, the cohort sought was the minimum amount for exposure and non-exposure groups or exposed and unexposed groups. If what is used is proportion data then for khohor research the p0 value in the formula above as the proportion of the sick in the unexposed population and p1 is the proportion that sakit in the exposed population or the value of p1 = p0 x RR (Relative Risk).

If the p-value is a continuous data (for example, average weight, height, BMI and so on) or not in proportional form, then the determination of the sample size for the group is carried out based on the following formula:

$$n = \frac{2(Z_{1-\alpha/2} + Z_{1-\beta})^2 \sigma^2}{(U1 - U2)^2}$$

Keterangan :

- n = jumlah sampel tiap kelompok
- $Z_{1-\alpha/2}$ = nilai pada distribusi normal standar yang sama d kemaknaan α (untuk α= 0,05 adalah 1,96)
- $Z_{1-\beta}$ = nilai pada distribusi normal standar yang sama (*power*) sebesar diinginkan (untuk ß=0,10 adalah 1,28
- σ = standar deviasi kesudahan (*outcome*)
- U1 = *mean outcome* kelompok tidak terpapar
- U2 = *mean outcome* kelompok terpapar

Case Control and Cohort Sample Formula 2

For example, let's say we want to find a minimal sample in a study on the effect of exclusive breastfeeding on the baby's weight. Using a meaningfulness level of 95% or *Alfa* = 0.05, and a power level of 90% or ß = 0.10, and the outcome observed is the baby's body weight set to have an assumption value of SD = 0.94 kg, and the estimated difference between the mean end value (outcome) of the weight of the unexposed group and the group exposed during the first 4 months of the baby's life (U0 - U1) of 0.6 kg (referring to the results of the Piwoz study, et al. 1994), then the estimated minimum number of samples needed by each observation group, whether exposed or not exposed are:

$$n = \frac{2(1{,}96 + 1{,}28)^2 \ (0{,}94)^2}{(0{,}6)^2}$$

Sample Count Example of Case Control and Cohort

= 51.5 people or rounded: 52 people/group

In khohor studies, it must be supplemented by the number of lost to follow or akalepas during the observation, usually assumed to be 15 %. In the example above, the minimum sample required becomes n= 52 (1 +0.15) = 59.8 babies or rounded to as many as 60 babies for each group either exposed or unexposed group or a total of 120 babies for both groups. Experimental Research According to Supranto J (2000) for experimental research with a complete randomized design, group random or factorial, it can simply be formulated:

(t-1) (r-1) > 15 where : t = number of treatment groups

j = number of replications For example: If there are 4 treatments, then the number of tests for each treatment can be calculated:

(4 -1) (r-1) > 15

(r-1) > 3/15

r > 6

5. Conclusion

The goal of most scientific research is to make general statements based on typical and limited observations. Basically, inference statistics allow researchers to make limited statements about the characteristics of a population, based on data from samples. The usefulness of the sample is very large meaning that in the sense that cost, labor and time can be saved, information can be obtained faster, obtaining data covering many properties, allowing the measurement of the reliability of estimates based on the results of the sample. Hopefully, the sampling method proposed can be useful for those who are interested in the field of social research

6. References

1. Anggarini, S. (n.d.). Sample Plan (Sampling design). *FKM-UI1979*.
2. Bhisma-Murti. (1997). *Principles and Methods of Epidemiological Research*. Gadjah Mata University Press.
3. GW, S., & Cochran WG. (1967). *Statistical Methods 6th ed*. Iowa State University Press.
4. Lemeshow, S., & David W.H.Jr. (1997). *Large Sample in Health Research*. Gadjahmada University Press.
5. Sujana. (1972). *Statistical Methods*. Tarsito.
6. Suparmoko, M. (1987). *Practical research methods (For the social and economic sciences)*. BPFE UNIV.

GAMA.

7. Supranto, J. (2000). *Sampling Techniques for Surveys and Experiments*. PT Rineka Cipta.