

## **DETERMINATION OF THE SAMPLE SIZE AND THE NUMBER OF FOLLOW-UP TIMES BY USING LINEAR PROGRAMMING**

**A. A. E. AHMED**

College of Business Administration  
Taibah University  
Medina  
Saudi Arabia

Helwan University  
Egypt  
e-mail: [ahmed63feb@gmail.com](mailto:ahmed63feb@gmail.com)

### **Abstract**

Researchers who are interested in follow-up studies face the problem of determining both sample size and follow-up length, and this paper investigates to deal with this problem. Mathematical programming is used to determine both sample size and follow-up length according to pre-determined cost.

### **1. Introduction**

There are two types of sampling. First, the type of sample data which are needed at one fixed point of time. Second, the type of sample data which are needed at sequence points of time. For the first type of sampling, the theory of sampling provides several methods of selecting a random sample. The second type of sampling includes the follow-up

---

2010 Mathematics Subject Classification: 62D99, 62P99, 90C05.

Keywords and phrases: follow-up studies, sampling, linear programming.

Received August 9, 2015

sample. Usually, the follow-up sample is followed for a period of time that is divided into small intervals. The sample follow-up data is recorded at each interval. New entrances and loss to follow-up are possible, so the sample may decrease or increase at each follow-up time, also the sample units may change during the study period.

## 2. Collecting Data in Follow-up Studies

Everitt [2] stated that there are two methods to collect data in follow-up studies:

### **First: Retrospective method**

“Retrospective” is a general term for studies in which all the events of interest occur prior to the onset of the study and findings are based on looking backward in time. He stated that this kind of studies has some advantages, among them:

- (1) They require fewer subjects than prospective studies.
- (2) They can be performed more quickly.
- (3) The investigator can perform some case-control studies by examining existing records.

This method is based on recorded data about units of study. Such allow researcher to follow the study-unit backward from the present time to a certain time-point in the past.

Several sources can provide this kind of data, such as police offices, insurance companies, patients’ records in hospitals, the personal in the governmental administrations... etc. Retrospective method has the following advantages:

- (1) Easiness of application.
- (2) Saving cost and time.
- (3) It does not require facing the study-units.

(4) Sometimes it is the only method to collect data, specially when the study sample is not available at the time of the study, or in the study of rare cases (or rare diseases).

Data collected in this way have the following disadvantages:

(1) It lacks accuracy and verification, because data are collected without facing the study units. For example, in medical follow-up studies, data can be collected via any individual who is a relative of the followed patient.

(2) Usually data is not enough to study.

### **Second: Prospective method**

In this method, study units are followed-up over a period of time. Researcher can design a data-base and collect data directly from study units. A researcher determines data needed to study and the method of its collection. Several tools can be used for data collection, such as telephone, questionnaire, interview, and mail. Usually, data collected in this method are enough to study. Prospective method needs more cost and time than retrospective method. Data obtained by this method are more accurate and verity than retrospective method. Sometimes prospective method is the only method to collect data, especially if the historical data of the study units are not found.

Finally, researcher can mix both retrospective and prospective method in follow-up studies, especially if the study units have a recorded follow-up data and still under follow-up and will be followed in the future. This method gives all advantages of both retrospective method and prospective method in addition to a longer period of follow-up.

### **3. Design of Follow-up Sample**

In follow-up studies, the researcher can only control the follow-up period and the follow-up interval. The follow-up period differs according to the nature of the studied system. It must be allow the system to develop, then transition probabilities (or any indicators or statistics) can be calculated.

The follow-up length may be either the same for the whole or differs according to some factors. After determining the follow up length, it is divided into many intervals. Researchers can determine either the number of follow-up times or the length of it. Follow-up data is recorded at each interval. Like follow-up length, follow-up intervals must be appropriate to the nature of the studied system; it may either be fixed for all followed units or differs according to their characteristics. The more the faster the development of the system the more the shorter the interval.

#### 4. Termination of Follow-up Study

Elandt-Johnson and Johnson [9] stated that in the follow-up studies, termination of observation may be controlled in many ways. The two most common termination methods are affected by the following schemes:

**Scheme 1.** Terminate observation at a preassigned time point (date), the resultant data are termed truncated (or type 1 censoring).

**Scheme 2.** Terminate observation becomes apparent when a preassigned number of failures have occur. The data are called censored (or type 2 censoring). In Scheme 1, the number of failures is random variable, whereas in Scheme 2, the time of termination is a random variable.

In summary, truncation and censoring are the terms used to indicated whether a fixed calendar date or a fixed number of failures has been used for termination of the study.

#### 5. Sample Size Determination in Follow-up Studies

Many researchers discussed how to determine the sample size in the follow-up studies. Sheps and Moustafa [7] discussed how to choose a desirable combination of sample size ( $n$ ) and duration of follow-up ( $t$ ) in the study of conception risk.

Palta and Mchugh [8] applied competing risk theory to the problem of sample size determination. El-Helbawy et al. [1] discussed the relationship between the sample size and length of the follow-up period for cohort studies with covariates.

## 6. Review

Grossman [5] suggested a strategy of designing the sampling in social experiments. His strategy based on preliminary tests of significance and it is a compromise between a large supposedly unpoolable sample and a smaller supposedly poolable sample. The author introduced an application of his suggested strategy.

Pezeshk [4] introduced a review of the subject of the determination of sample size in designing medical studies. The author introduced a number of papers that covered Bayesian methods of sample size determination. The results of the fully Bayesian sample size determination to an actual clinical trial were discussed.

Naing et al. [6] threw a light on the problem of determination the sample size for a study estimating a population prevalence. They introduced simple formulas to calculate the sample size in various situations. They discussed the sentence “The larger the sample size the better the study” is not always true.

Teddle and Yu [3] discussed the mixed methods (MM) sampling techniques. They introduced four MM sampling prototypes: basic MM sampling strategies, sequential MM sampling, concurrent MM sampling, and multilevel MM sampling. They divided the sampling techniques for the social and behavioural sciences into four categories:

### I. Probability sampling

- A. Random sampling.
- B. Stratified sampling.
- C. Cluster sampling.
- D. Sampling using multiple probability techniques.

**II. Purposive sampling**

- A. Sampling to achieve representative or comparability.
- B. Sampling special or unique cases.
- C. Sequential sampling.
- D. Sampling using multiple purposive techniques.

**III. Convenience sampling**

- A. Captive sample.
- B. Volunteer sample.

**IV. Mixed methods sampling**

- A. Basic mixed methods sampling.
- B. Sequential mixed methods sampling.
- C. Concurrent mixed methods sampling.
- D. Multilevel mixed methods sampling.
- E. Combination of mixed methods sampling strategies.

**7. Determination of the Sample Size and the Number of Follow-up Times by Using Linear Programming****Assumptions**

- The follow-up sample size  $n$  is the summation of follow-up times of all follow-up units, and its lower limit  $n_1$ .
- In follow-up process new entrance, losses to follow-up and withdrawals are allowed.
- In the follow-up process each follow-up time contains a number of follow-up units. It means that there are  $k$  follow-up-group times with minimum value  $k_1$ .

- Each follow-up-group time requires a fixed cost  $c_f$ . The variable cost of each follow-up time is  $c_v$ .

- $TC$  is the total cost of follow-up process with a maximum  $TC_u$ .

Under the above assumptions, we can form the problem of determination the sample size  $n$  and the number of group-follow-up times  $k$ , then the follow-up length as a problem of minimization the total cost of follow-up process as follows:

$$\text{Minimize: } TC = c_f k + c_v n$$

Subject to:

$$TC \leq TC_u$$

$$k \geq k_1$$

$$n \geq n_1$$

$$k, n > 0.$$

This is an initial formula, it is flexible and can be modified according to the nature of the actual problem. More constraints can be added according to the relationship between the variables of the problem.

We can use this approach to determinate the sample size  $n$  and the number of group-follow-up times  $k$ , then the follow-up length in many fields, such as:

(1) Health care field.

(2) Maintenance of the machines in industrial enterprises systems.

(3) Any system that needs to follow-up both at the level of individuals or institutions.

### A simulation case

An organization has units that are followed periodically.  $n$  is the number of follow-up times at a year with minimum  $n_1 = 40$  units. There are  $k$  group-follow-up times with minimum value  $k_1 = 12$ . Each follow-up time requires a fixed cost  $c_f = \$100$ . Each follow-up time needs a variable cost  $c_v = \$20$ .  $TC$  is the total cost of follow-up process with a maximum  $TC_u = \$10000$ . Then, we can form the problem of determination the group-follow-up times  $k$  and the follow-up sample  $n$  as follows:

$$\text{Minimize: } TC = 100k + 20n$$

Subject to:

$$100k + 20n \leq 10000$$

$$k \geq 12$$

$$n \geq 40$$

$$k, n > 0.$$

By solving this problem, we find three feasible solutions at the following three points (a) (12, 40), (b) (92, 40), and (c) (12, 440). The total cost at (a) equal \$2000, at (b) equal \$10000, and at (c) equal \$10000. The optimal solution at the point (a) with a total cost = \$2000.

### References

- [1] A. El-Helbawy, M. Adel-Sabour and S. M. Abdel-Nabi, Sample size determination for follow-up studies of lifetime with covariates, The first conference on statistics and its commercial and economical applications, Helwan University, Faculty of Commerce and Business Administration, Dept. of Mathematics, Insurance, and Applied Statistics (2002), 123-138.
- [2] B. S. Everitt, The Cambridge Dictionary of Statistics, Cambridge University Press, 1998.
- [3] C. Teddlie and F. Yu, Mixed methods sampling a typology with examples, Journal of Mixed Methods Research 1 (2007), 77-100.



- [4] H. Pezeshk, Bayesian techniques for sample size determination in clinical trials: A short review, *Statistical Methods in Medical Research* 12 (2003), 489-504.
- [5] J. B. Grossman, Optimal sample designs with preliminary tests of significance, *Journal of Business & Economics Statistics* 4(2) (1986), 76-171.
- [6] L. Naing, T. Winn and B. N. Rusli, Practical Issues in Calculating the Sample Size of Prevalence Studies, *Archives of Orofacial Sciences*, 2006.
- [7] M. C. Sheps and A. F. M. Moustafa, Some considerations affection the design of follow-up studies, *Egypt. Statist. J.* 19(2) (1975), 76-85.
- [8] M. Palta and R. Mchugh, Adjusting for losses to follow-up in sample size determination for cohort studies, *J. Chiron. Dis.* 32 (1979), 315-326.
- [9] R. C. Elandt-Johnson and N. L. Johnson, *Survival Models and Data Analysis*, John Wiley, 1980.

